

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: August 2, 2004, 14:35:42 ; Search time 44.9 Seconds

(without alignments)  
950.215 Million cell updates/sec

Title: US-09-543-407-12

Perfect score: 779

Sequence: 1 MKLLKVAAPAAIIVSGSALA.....VTRVVTHEAAHANNATANQY 151

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1586107 seqs, 282547505 residues

Total number of hits satisfying chosen parameters: 1586107

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%  
Listing first 45 summaries

Database : A Geneseq\_29Jan04:\*

- 1: geneseqp1980s:\*
- 2: geneseqp1980s:\*
- 3: geneseqp2000s:\*
- 4: geneseqp2001s:\*
- 5: geneseqp2002s:\*
- 6: geneseqp2003as:\*
- 7: geneseqp2003bs:\*
- 8: geneseqp2004s:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	% Match	Length	DB ID	Description
1	779	100.0	151	3 AAB36346	Aab36346 Agfa::PT3
2	696	89.3	151	3 AAB36347	Aab36347 Agfa::PT3
3	695	89.2	151	2 AAR74625	Aar74625 Agfa sequ
4	695	89.2	151	3 AAB36341	Aab36341 Salmonell
5	690	88.6	151	2 AAW23570	Aaw23570 Salmonell
6	655	84.1	151	3 AAB36352	Aab36352 Agfa::PT3
7	614	78.8	151	3 AAB36353	Aab36353 Agfa::PT3
8	613	78.7	151	3 AAB36349	Aab36349 Agfa::PT3
9	611	78.4	151	3 AAB36350	Aab36350 Agfa::PT3
10	605	77.7	151	3 AAB36354	Aab36354 Agfa::PT3
11	604	77.5	151	3 AAB36351	Aab36351 Agfa::PT3
12	603	77.4	151	3 AAB36355	Aab36355 Agfa::PT3
13	580	74.5	151	3 AAB36348	Aab36348 Agfa::PT3
14	560	71.9	120	2 AAR62761	Aar62761 Agfa sequ
15	560	71.9	120	2 AAW23569	Aaw23569 Salmonell
16	520	66.8	151	3 AAB36343	Aab36343 Escherich
17	515	66.1	151	7 ABR82651	Abr82651 E. coli C
18	485	62.3	142	2 AAR52664	Aar52664 Fibronect
19	413	53.0	122	2 AAR52663	Aar52663 FNE curli
20	237	30.4	45	3 AAB36316	Aab36316 Salmonell
21	132	16.9	22	3 AAB36318	Aab36318 Salmonell
22	123	15.8	23	3 AAB36321	Aab36321 Salmonell
23	123	15.8	23	3 AAB36326	Aab36326 Salmonell
24	123	15.8	23	3 AAB36338	Aab36338 Salmonell
25	115	14.8	22	3 AAB36325	Aab36325 Salmonell

## ALIGNMENTS

## RESULT 1

AAB36346

ID AAB36346 standard; protein; 151 AA.

XX

AC AAB36346;

XX

DT 26-FEB-2001 (first entry)

XX

DE Agfa::PT3#1 amino acid sequence SEQ ID NO:12.

XX

XX Salmonella; agfa; chromosomal gene replacement; fimbrin; epitope;

KW vaccine; immune response; immunogen.

XX

OS Salmonella enteritidis.

OS

OS Escherichia coli.

OS

XX Synthetic.

XX

FN WO200060102-A2.

XX

PD 12-OCT-2000.

XX

PF 05-APR-2000; 2000WO-CA000356.

XX

PR 05-APR-1999; 99US-0127888P.

XX

PA Aab36347 Agfa::PT3

XX

PI (UYVI-) UNIV VICTORIA.

XX

PI White AP, Doran JL, Collison SK, Kay WW;

XX

DR WPI: 2000-672631/65.

XX

DR N-PSDB; AAC64622.

XX

PT Recombinant agfa gene having a segment replaced by a foreign DNA sequence

XX

PT which encodes foreign epitope or antigen, expresses recombinant Agfa

XX

PS protein useful for eliciting immune response in animal.

XX

PS Disclosure; Page 135; 139pp; English.

XX

CC The present invention describes a recombinant agfa gene (I) where a

CC

CC segment of the gene has been replaced by a segment of a foreign DNA

CC

CC sequence which encodes a foreign epitope or antigen. Also described are:

CC

CC (1) use of thin aggregative fimbriae (SFP17/TAF) nucleation depended

CC

CC assembly system of strains of Salmonella, Escherichia coli and

CC

CC Enterobacteriaceae for the production of fimbriae comprising recombinant

CC

CC Agfa, CsgA and Agfa-homologue fimbria subunits, respectively; (2)

CC

CC directing recombination of a recombinant gene into the chromosome of the

CC

CC homologous species; (3) directing recombination of a recombinant gene

CC

CC back into the chromosome of the homologous species, replacing the native

CC copy of that gene; and (4) eliciting an immune response in an animal,  
 CC comprising separating an amino acid polymer comprising a recombinant Agfa  
 CC protein containing a replacement segment or segments of foreign amino  
 CC acid sequence or sequences grown on a Salmonella, E. coli or  
 CC Enterobacteriaceae host cell, from the host cell, and introducing the  
 CC polymer into the animal in conjunction with a carrier or diluent. (1) is  
 CC useful for the expression of recombinant Agfa protein which is useful for  
 CC eliciting an immune response in an animal. In a fimbrial presentation  
 CC system the heterologous antigens are presented in high numbers (up to  
 CC 500,000 copies/cell), the hybrid fimbrial protein possesses both the  
 CC immunogenicity and adhesion properties relevant for an efficient live  
 CC vaccine, the carrier fimbrial subunit proteins are usually strong  
 CC immunogens, which may be important for directing an immune response  
 CC against the inserted epitope, and hybrid fimbriae are easy and  
 CC inexpensive to purify in large amount. The present sequence is given in  
 CC the exemplification of the present invention  
 XX  
 SQ Sequence 151 AA;

Query Match 100.0%; Score 779; DB 3; Length 151;  
 Best Local Similarity 100.0%; Pred. No. 2e-66;  
 Matches 151; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 MKLLKVAFAAIVVSGSALAGVVPQWGGGNGHNGGNSGPDSTLSIYQYGSANAALALQ 60  
 DB 1 MKLLKVAFAAIVVSGSALAGVVPQWGGGNGHNGGNSGPDSTLSIYQYGSANAALALQ 60  
 QY 61 SDARKSETTITQSGYNGADVGQADNSTIELTQNGFRNNATIDQWNAKNSDITVGYGG 120  
 DB 61 SDARKSETTITQSGYNGADVGQADNSTIELTQNGFRNNATIDQWNAKNSDITVGYGG 120  
 QY 121 NNAALVNYDQLVTRVTHEMAHANATANY 151  
 DB 121 NNAALVNYDQLVTRVTHEMAHANATANY 151

RESULT 2  
 AAB36347  
 ID AAB36347 standard; protein; 151 AA.  
 XX  
 AC AAB36347;  
 XX  
 DT 26-FEB-2001 (first entry)  
 XX  
 DE Agfa::PT3#2 amino acid sequence SEQ ID NO:14.  
 XX  
 KW Salmonella; agfa; chromosomal gene replacement; fimbria; epitope;  
 KW vaccine; immune response; immunogen.  
 XX  
 OS Salmonella enteritidis.  
 OS Escherichia coli.  
 OS Synthetic.  
 XX  
 PN W0200060102-A2.  
 XX  
 PD 12-OCT-2000.  
 XX  
 XX 05-APR-2000; 2000WO-CA000356.  
 XX  
 XX 05-APR-1999; 99US-0127888P.  
 XX  
 PA (UYVI-) UNIV VICTORIA.  
 XX  
 XX White AP, Doran JL, Collison SK, Kay MW;  
 XX  
 XX WPI: 2000-672631/65.  
 DR N-PSDB; AAC64623.  
 XX  
 XX Recombinant agfa gene having a segment replaced by a foreign DNA sequence  
 PT which encodes foreign epitope or antigen, expresses recombinant Agfa  
 PT protein useful for eliciting immune response in animal.  
 XX  
 PS Disclosure; Page 136; 139pp; English.

XX The present invention describes a recombinant agfa gene (I) where a  
 CC segment of the gene has been replaced by a segment of a foreign DNA  
 CC sequence which encodes a foreign epitope or antigen. Also described are:  
 CC (1) use of thin aggregative fimbriae (SEF17/TAF) nucleation depended  
 CC assembly system of strains of Salmonella, Escherichia coli and  
 CC Enterobacteriaceae for the production of fimbriae comprising recombinant  
 CC Agfa, CsgA and Agfa-homologue fimbrial subunits, respectively; (2)  
 CC directing recombination of a recombinant gene into the chromosome of the  
 CC homologous species; (3) directing recombination of a recombinant gene  
 CC back into the chromosome of the homologous species, replacing the native  
 CC copy of that gene; and (4) eliciting an immune response in an animal,  
 CC comprising separating an amino acid polymer comprising a recombinant Agfa  
 CC protein containing a replacement segment or segments of foreign amino  
 CC acid sequence or sequences grown on a Salmonella, E. coli or  
 CC Enterobacteriaceae host cell, from the host cell and introducing the  
 CC polymer into the animal in conjunction with a carrier or diluent. (1) is  
 CC useful for the expression of recombinant Agfa protein which is useful for  
 CC eliciting an immune response in an animal. In a fimbrial presentation  
 CC system the heterologous antigens are presented in high numbers (up to  
 CC 500,000 copies/cell), the hybrid fimbrial protein possesses both the  
 CC immunogenicity and adhesion properties relevant for an efficient live  
 CC vaccine, the carrier fimbrial subunit proteins are usually strong  
 CC immunogens, which may be important for directing an immune response  
 CC against the inserted epitope, and hybrid fimbriae are easy and  
 CC inexpensive to purify in large amount. The present sequence is given in  
 CC the exemplification of the present invention  
 XX  
 SQ Sequence 151 AA;

Query Match 89.3%; Score 696; DB 3; Length 151;  
 Best Local Similarity 87.6%; Pred. No. 1.7e-59;  
 Matches 141; Conservative 0; Mismatches 0; Indels 20; Gaps 2;  
 QY 1 MKLLKVAFAAIVVSGSALAGVVPQWGGGNGHNGGNSGPDSTLSIYQYGSANAALALQ 60  
 DB 1 MKLLKVAFAAIVVSGSALAGVVPQWGGGNGHNGGNSGPDSTLSIYQYGSANAALALQ 60  
 QY 61 SDARKSETTITQSGYNGADVGQADNSTIELTQNGFRNNATIDQWNAKNSDITVGYGG 120  
 DB 61 SDARKSETTITQSGYNGADVGQADNSTIELTQNGFRNNATIDQWNAKNSDITVGYGG 117  
 QY 121 NNAALVNYDQLVTRVTHEMAHA-----NNATANQY 151  
 DB 118 -----YDQLVTRVTHEMAHASVWVRQVGFNNATANQY 151

RESULT 3  
 AAR74625  
 ID AAR74625 standard; protein; 151 AA.  
 XX  
 AC AAR74625;  
 XX  
 DT 25-MAR-2003 (revised)  
 DT 26-JUN-1995 (first entry)  
 XX  
 DE Agfa sequence.  
 XX  
 KW Salmonella; Agfa; vaccine.  
 XX  
 OS Salmonella.  
 XX  
 PN W09425598-A2.  
 XX  
 PD 10-NOV-1994.  
 XX  
 PF 26-APR-1994; 94WO-IB000207.  
 XX  
 PR 26-APR-1993; 93US-00054452.  
 XX  
 PA (UYVI-) UNIV VICTORIA INNOVATION & DEV CORP.  
 PA (KING/) KING J.  
 XX



PI Kay WW, Collinson SK, Clouthier SC, Doran JL;  
 XX WPI; 1994-358275/44.  
 DR N-PSDB; AAQ87467.  
 XX  
 PT Eliciting an immune response to Salmonella - using attenuated Salmonella  
 PT strains, vector constructs, or compsns. contg. fimbrial type proteins.  
 XX  
 PS Disclosure; Fig 7B; 95pp; English.  
 XX  
 CC The Salmonella Agfa protein and DNA are used in vaccine and genetic  
 CC immunization compositions, respectively, to elicit an immune response to  
 CC Salmonella in animals (e.g. food producing animals) and humans. (Updated  
 CC on 25-MAR-2003 to correct PN field.)  
 XX  
 SQ Sequence 151 AA;  
 Query Match 89.2%; Score 695; DB 2; Length 151;  
 Best Local Similarity 90.1%; Pred. No. 2.2e-58;  
 Matches 136; Conservative 4; Mismatches 11; Indels 0; Gaps 0;  
 QY 1 MKLLKVAFAAIVVSGSALAGVVPQWGGGNGHNGGNSGPDSTLSIYQYGSANAALALQ 60  
 DB 1 MKLLKVAFAAIVVSGSALAGVVPQWGGGNGHNGGNSGPDSTLSIYQYGSANAALALQ 60  
 QY 61 SDARKSETTITQSGYNGADVGGADNSTIELTONGFRNNATIDQWNAKNSDITVGYGG 120  
 DB 61 SDARKSETTITQSGYNGADVGGADNSTIELTONGFRNNATIDQWNAKNSDITVGYGG 120  
 QY 121 NNAALVNYDQLVTRVVTHEMAHANNATANQY 151  
 DB 121 NNAALVNYDQLVTRVVTHEMAHANNATANQY 151  
 RESULT 4  
 AAB36341  
 ID AAB36341 standard; protein; 151 AA.  
 AC AAB36341;  
 XX  
 DT 26-FEB-2001 (first entry)  
 DE Salmonella enteritidis Agfa amino acid sequence SEQ ID NO:5.  
 DE  
 DE Salmonella; agfa; chromosomal gene replacement; fimbria; epitope;  
 KW vaccine; immune response; immunogen.  
 XX  
 OS Salmonella enteritidis.  
 XX  
 PN WO200060102-A2.  
 XX  
 PD 12-OCT-2000.  
 XX  
 PF 05-APR-2000; 2000WO-CA000356.  
 XX  
 PR 05-APR-1999; 99US-0127888P.  
 XX  
 PA (UYVI-) UNIV VICTORIA.  
 XX  
 PI White AP, Doran JL, Collinson SK, Kay WW;  
 XX  
 DR WPI; 2000-672631/65.  
 DR N-PSDB; AAC64617.  
 XX  
 CC Recombinant agfa gene having a segment replaced by a foreign DNA sequence  
 PT which encodes foreign epitope or antigen, expresses recombinant Agfa  
 PT protein useful for eliciting immune response in animal.  
 XX  
 PS Disclosure; Page 135; 139pp; English.  
 XX  
 CC The present invention describes a recombinant agfa gene (I) where a  
 CC segment of the gene has been replaced by a segment of a foreign DNA  
 CC sequence which encodes a foreign epitope or antigen. Also described are:

CC (1) use of thin aggregative fimbriae (SEF17/TAF) nucleation depended  
 CC assembly system of strains of Salmonella, Escherichia coli and  
 CC Enterobacteriaceae for the production of fimbriae comprising recombinant  
 CC Agfa, CsgA and Agfa-homologue fimbrial subunits, respectively; (2)  
 CC directing recombination of a recombinant gene into the chromosome of the  
 CC homologous species; (3) directing recombination of a recombinant gene  
 CC back into the chromosome of the homologous species, replacing the native  
 CC copy of that gene; and (4) eliciting an immune response in an animal  
 CC comprising separating an amino acid polymer comprising a recombinant Agfa  
 CC protein containing a replacement segment or segments of foreign amino  
 CC acid sequence or sequences grown on a Salmonella, E. coli or  
 CC Enterobacteriaceae host cell, from the host cell and introducing the  
 CC polymer into the animal in conjunction with a carrier or diluent. (I) is  
 CC useful for the expression of recombinant Agfa protein which is useful for  
 CC eliciting an immune response in an animal. In a fimbrial presentation  
 CC system the heterologous antigens are presented in high numbers (up to  
 CC 500,000 copies/cell), the hybrid fimbrial protein possesses both the  
 CC immunogenicity and adhesion properties relevant for an efficient live  
 CC vaccine, the carrier fimbrial subunit proteins are usually strong  
 CC immunogens, which may be important for directing an immune response  
 CC against the inserted epitope, and hybrid fimbriae are easy and  
 CC inexpensive to purify in large amount. The present sequence is given in  
 CC the exemplification of the present invention  
 XX  
 SQ Sequence 151 AA;  
 Query Match 89.2%; Score 695; DB 3; Length 151;  
 Best Local Similarity 90.1%; Pred. No. 2.2e-58;  
 Matches 136; Conservative 4; Mismatches 11; Indels 0; Gaps 0;  
 QY 1 MKLLKVAFAAIVVSGSALAGVVPQWGGGNGHNGGNSGPDSTLSIYQYGSANAALALQ 60  
 DB 1 MKLLKVAFAAIVVSGSALAGVVPQWGGGNGHNGGNSGPDSTLSIYQYGSANAALALQ 60  
 QY 61 SDARKSETTITQSGYNGADVGGADNSTIELTONGFRNNATIDQWNAKNSDITVGYGG 120  
 DB 61 SDARKSETTITQSGYNGADVGGADNSTIELTONGFRNNATIDQWNAKNSDITVGYGG 120  
 QY 121 NNAALVNYDQLVTRVVTHEMAHANNATANQY 151  
 DB 121 NNAALVNYDQLVTRVVTHEMAHANNATANQY 151  
 RESULT 5  
 AAW23570  
 ID AAW23570 standard; protein; 151 AA.  
 XX  
 AC AAW23570;  
 XX  
 DT 25-MAR-2003 (revised)  
 DT 29-SEP-1997 (first entry)  
 DE  
 DE Salmonella enteritidis 27655-3b agfa.  
 XX  
 KW Enteropathogenic bacteria; enterobacteria; S. enteritidis; antibody.  
 XX  
 OS Salmonella enteritidis.  
 XX  
 FH Key Location/Qualifiers  
 FT Misc-difference 123  
 FT /note= "Encoded by GCC"  
 XX  
 PN US5635617-A.  
 XX  
 PD 03-JUN-1997.  
 XX  
 PF 26-APR-1994; 94US-00233788.  
 XX  
 PR 26-APR-1993; 93US-00054452.  
 XX  
 PA (UYVI-) UNIV VICTORIA INNOVATION & DEV CORP.  
 XX Collinson SK, Kay WW, Doran JL;  
 PI

```

XX WPI; 1997-309886/28.
XX N-PSDB; AAT74142.
XX
XX Isolated Salmonella gene agfa - used for diagnosis of Salmonella or
XX enteropathogenic bacteria of the Enterobacteria family.
XX
XX Example 2; Fig 7; 85pp; English.
XX
XX The present sequence represents agfa encoded by the full agfa gene
XX derived from Salmonella enteritidis 27655-3b. The nucleic acid can be
XX used to provide diagnostic assays for Salmonella and/or enteropathogenic
XX bacteria of the family Enterobacteria. It can also be used to provide
XX proteins and antibodies which can be used for assays. The nucleic acid
XX sequence can be used to provide probes or primers which can specifically
XX hybridise to nucleic acid molecules from greater than 99% of Salmonella
XX strains that are pathogenic to warm-blooded animals relative to nucleic
XX acid molecules from virtually all other microbial organisms. (Updated on
XX 25-MAR-2003 to correct PF field.)
XX
XX Sequence 151 AA;
XX
XX Query Match      88.6%; Score 690; DB 2; Length 151;
XX Best Local Similarity 89.4%; Pred. No. 6.5e-58;
XX Matches 135; Conservative 4; Mismatches 12; Indels 0; Gaps 0;
XX
XX QY 1 MKLLKVAFAAIVVSGSALAGVVPQWGGGNGHNGGNSGPDSTLSIYQYGSANAALALQ 60
XX DB 1 MKLLKVAFAAIVVSGSALAGVVPQWGGGNGHNGGNSGPDSTLSIYQYGSANAALALQ 60
XX
XX QY 61 SDARKSETTITQSGYNGADVGQADNSTIELTQNGFRNNATIDQWNAKNSDITVGYGG 120
XX DB 61 SDARKSETTITQSGYNGADVGQADNSTIELTQNGFRNNATIDQWNAKNSDITVGYGG 120
XX
XX QY 121 NNAALVNDQLVTRVVTHEMAHANNATANY 151
XX DB 121 NNPALVNTASDSVWVQVFGNNATANY 151
XX
XX RESULT 6
XX AAB36352
XX ID AAB36352 standard; protein; 151 AA.
XX AC AAB36352;
XX XX 26-FEB-2001 (first entry)
XX DT Agfa::PT3#7 amino acid sequence SEQ ID NO:24.
XX DE Salmonella; agfa; chromosomal gene replacement; fimbrin; epitope;
XX KW vaccine; immune response; immunogen.
XX OS Salmonella enteritidis.
XX OS Escherichia coli.
XX OS Synthetic.
XX XX WO2000060102-A2.
XX XX 12-OCT-2000.
XX XX 05-APR-2000; 2000WO-CA000356.
XX XX 05-APR-1999; 99US-0127888P.
XX XX (UUVI-) UNIV VICTORIA.
XX XX White AP, Doran JL, Collison SK, Kay WW;
XX XX WPI; 2000-672631/65.
XX XX N-PSDB; AAC64628.
XX XX Recombinant agfa gene having a segment replaced by a foreign DNA sequence
XX PT which encodes foreign epitope or antigen, expresses recombinant Agfa

```

protein useful for eliciting immune response in animal.

Dislosure; Page 138; 139pp; English.

The present invention describes a recombinant agfa gene (I) where a segment of the gene has been replaced by a segment of a foreign DNA sequence which encodes a foreign epitope or antigen. Also described are: (1) use of thin aggregative fimbriae (SEF17/TAF) nucleation depended assembly system of strains of Salmonella, Escherichia coli and Enterobacteriaceae for the production of fimbriae comprising recombinant Agfa, CsgA and agfa-homologue fimbrin subunits, respectively; (2) directing recombination of a recombinant gene into the chromosome of the homologous species; (3) directing recombination of a recombinant gene back into the chromosome of the homologous species, replacing the native copy of that gene; and (4) eliciting an immune response in an animal, comprising separating an amino acid polymer comprising a recombinant Agfa protein containing a replacement segment or segments of foreign amino acid sequence or sequences grown on a Salmonella, E. coli or Enterobacteriaceae host cell, from the host cell and introducing the polymer into the animal in conjunction with a carrier or diluent. (I) is useful for the expression of recombinant Agfa protein which is useful for eliciting an immune response in an animal. In a fimbrial presentation system the heterologous antigens are presented in high numbers (up to 500,000 copies/cell), the hybrid fimbrin protein possesses both the immunogenicity and adhesion properties relevant for an efficient live vaccine, the carrier fimbrial subunit proteins are usually strong immunogens, which may be important for directing an immune response against the inserted epitope, and hybrid fimbriae are easy and inexpensive to purify in large amount. The present sequence is given in the exemplification of the present invention

Sequence 151 AA;

Query Match 84.1%; Score 655; DB 3; Length 151;

Best Local Similarity 79.8%; Pred. No. 1.4e-54;

Matches 134; Conservative 0; Mismatches 0; Indels 34; Gaps 2;

QY 1 MKLLKVAFAAIVVSGSALAGVVPQWGGGNGHNGGNSGPDSTLSIYQYGSANAALALQ 60

DB 1 MKLLKVAFAAIVVSGSALAGVVPQWGGGNGHNGGNSGPDSTLSIYQYGSANAALALQ 60

QY 61 SDARKSETTITQSGYNGADVGQADNSTIELTQNGFRNNATIDQWNAKNSDITVGYGG 120

DB 61 SDARKSETTITQSGYNGADVGQADNSTIELTQNGFRNNATIDQWNAKNSDITVGYGG 120

QY 121 NNAALVNDQLVTRVVTHEMAHA-----NNATANY 151

DB 110 -----NYDQLVTRVVTHEMAHANQTASDSVWVQVFGNNATANY 151

RESULT 7

AAB36353

ID AAB36353 standard; protein; 151 AA.

XX AC AAB36353;

XX XX 26-FEB-2001 (first entry)

XX DT Agfa::PT3#8 amino acid sequence SEQ ID NO:26.

XX DE Salmonella; agfa; chromosomal gene replacement; fimbrin; epitope;

XX KW vaccine; immune response; immunogen.

XX OS Salmonella enteritidis.

XX OS Escherichia coli.

XX OS Synthetic.

XX XX WO2000060102-A2.

XX XX 12-OCT-2000.

XX XX 05-APR-2000; 2000WO-CA000356.

XX XX

```

PR 05-APR-1999; 99US-0127888P.
PA (UYVI-) UNIV VICTORIA.
XX
PI White AP, Doran JL, Collison SK, Kay WW;
XX
XX WPI; 2000-672631/65.
XX N-PSDB; AAC64629.
XX
XX Recombinant agfa gene having a segment replaced by a foreign DNA sequence
PT which encodes foreign epitope or antigen, expresses recombinant Agfa
PT protein useful for eliciting immune response in animal.
XX
XX Disclosure; Page 138; 139pp; English.
XX
XX The present invention describes a recombinant agfa gene (I) where a
CC segment of the gene has been replaced by a segment of a foreign DNA
CC sequence which encodes a foreign epitope or antigen. Also described are:
CC (1) use of thin aggregative fimbriae (SEF17/TAf) nucleation depended
CC assembly system of strains of Salmonella, Escherichia coli and
CC Enterobacteriaceae for the production of fimbriae comprising recombinant
CC Agfa, CsgA and Agfa-homologue fimbrial subunits, respectively; (2)
CC directing recombination of a recombinant gene into the chromosome of the
CC homologous species; (3) directing recombination of a recombinant gene
CC back into the chromosome of the homologous species, replacing the native
CC copy of that gene; and (4) eliciting an immune response in an animal,
CC comprising separating an amino acid polymer comprising a recombinant Agfa
CC protein containing a replacement segment or segments of foreign amino
CC acid sequence or sequences grown on a Salmonella, E. coli or
CC Enterobacteriaceae host cell, from the host cell and introducing the
CC polymer into the animal in conjunction with a carrier or diluent. (I) is
CC useful for the expression of recombinant Agfa protein which is useful for
CC eliciting an immune response in an animal. In a fimbrial presentation
CC system the heterologous antigens are presented in high numbers (up to
CC 500,000 copies/cell), the hybrid fimbrial protein possesses both the
CC immunogenicity and adhesion properties relevant for an efficient live
CC vaccine, the carrier fimbrial subunit proteins are usually strong
CC immunogens, which may be important for directing an immune response
CC against the inserted epitope, and hybrid fimbriae are easy and
CC inexpensive to purify in large amount. The present sequence is given in
CC the exemplification of the present invention
XX
XX Sequence 151 AA;
XX
XX Query Match 78.8%; Score 614; DB 3; Length 151;
XX Best Local Similarity 80.8%; Pred. No. 1.2e-50;
XX Matches 122; Conservative 6; Mismatches 23; Indels 0; Gaps 0;
XX
QY 1 MKLLKVAAPAAIIVSGSALAGVVPQWGGGNGHNGSGSPDSTLSIYQYGSANALALQ 60
DB 1 MKLLKVAAPAAIIVSGSALAGVVPQWGGGNGHNGSGSPDSTLSIYQYGSANALYDQ 60
XX
QY 61 SDARKSETTITQSGYNGADVQGGADNSTIELTQNGFRNNATIDQWNAKNSDITVQYGG 120
DB 61 LVTRVVTHEMAHAGYNGADVQGGADNSTIELTQNGFRNNATIDQWNAKNSDITVQYGG 120
XX
QY 121 NNAALVNYDQLVTRVVTHEMAHANNATANYQ 151
DB 121 NNAALVNYDQLVTRVVTHEMAHANNATANYQ 151
XX
RESULT 8
AAB36349
ID AAB36349 standard; protein; 151 AA.
XX
AC AAB36349;
XX
XX 26-FEB-2001 (first entry)
XX
XX Agfa::PT3#4 amino acid sequence SEQ ID NO:18.
XX
XX Salmonella; agfa; chromosomal gene replacement; fimbrial; epitope;
XX vaccine; immune response; immunogen.
XX

```

```

XX Salmonella enteritidis.
XX Escherichia coli.
XX Synthetic.
XX
XX WO200060102-A2.
XX
XX 12-OCT-2000.
XX
XX 05-APR-2000; 2000WO-CA000356.
XX
XX 05-APR-1999; 99US-0127888P.
XX
XX (UYVI-) UNIV VICTORIA.
XX
XX White AP, Doran JL, Collison SK, Kay WW;
XX
XX WPI; 2000-672631/65.
XX N-PSDB; AAC64625.
XX
XX Recombinant agfa gene having a segment replaced by a foreign DNA sequence
PT which encodes foreign epitope or antigen, expresses recombinant Agfa
PT protein useful for eliciting immune response in animal.
XX
XX Disclosure; Page 136; 139pp; English.
XX
XX The present invention describes a recombinant agfa gene (I) where a
CC segment of the gene has been replaced by a segment of a foreign DNA
CC sequence which encodes a foreign epitope or antigen. Also described are:
CC (1) use of thin aggregative fimbriae (SEF17/TAf) nucleation depended
CC assembly system of strains of Salmonella, Escherichia coli and
CC Enterobacteriaceae for the production of fimbriae comprising recombinant
CC Agfa, CsgA and Agfa-homologue fimbrial subunits, respectively; (2)
CC directing recombination of a recombinant gene into the chromosome of the
CC homologous species; (3) directing recombination of a recombinant gene
CC back into the chromosome of the homologous species, replacing the native
CC copy of that gene; and (4) eliciting an immune response in an animal,
CC comprising separating an amino acid polymer comprising a recombinant Agfa
CC protein containing a replacement segment or segments of foreign amino
CC acid sequence or sequences grown on a Salmonella, E. coli or
CC Enterobacteriaceae host cell, from the host cell and introducing the
CC polymer into the animal in conjunction with a carrier or diluent. (I) is
CC useful for the expression of recombinant Agfa protein which is useful for
CC eliciting an immune response in an animal. In a fimbrial presentation
CC system the heterologous antigens are presented in high numbers (up to
CC 500,000 copies/cell), the hybrid fimbrial protein possesses both the
CC immunogenicity and adhesion properties relevant for an efficient live
CC vaccine, the carrier fimbrial subunit proteins are usually strong
CC immunogens, which may be important for directing an immune response
CC against the inserted epitope, and hybrid fimbriae are easy and
CC inexpensive to purify in large amount. The present sequence is given in
CC the exemplification of the present invention
XX
XX Sequence 151 AA;
XX
XX Query Match 78.7%; Score 613; DB 3; Length 151;
XX Best Local Similarity 80.8%; Pred. No. 1.5e-50;
XX Matches 122; Conservative 7; Mismatches 22; Indels 0; Gaps 0;
XX
QY 1 MKLLKVAAPAAIIVSGSALAGVVPQWGGGNGHNGSGSPDSTLSIYQYGSANALALQ 60
DB 1 MKLLKVAAPAAIIVSGSALAGVVPQWGGGNGHNGSGSPDSTLSIYQYGSANALALQ 60
XX
QY 61 SDARKSETTITQSGYNGADVQGGADNSTIELTQNGFRNNATIDQWNAKNSDITVQYGG 120
DB 61 SDARKSETTITQSGYNGADVQGGADNSTIELTQNGFRNNATIDQWNAKNSDITVQYGG 120
XX
QY 121 NNAALVNYDQLVTRVVTHEMAHANNATANYQ 151
DB 121 NNAALVNYDQLVTRVVTHEMAHANNATANYQ 151
XX
RESULT 9

```

AAB36350  
 ID AAB36350 standard; protein; 151 AA.  
 AC AAB36350;  
 XX  
 DT 26-FEB-2001 (first entry)  
 XX  
 DE Agfa::PT3#5 amino acid sequence SEQ ID NO:20.  
 XX  
 KW Salmonella; agfa; chromosomal gene replacement; fimbria; epitope;  
 KW vaccine; immune response; immunogen.  
 XX  
 OS Salmonella enteritidis.  
 OS Escherichia coli.  
 OS Synthetic.  
 XX  
 FN WO2000060102-A2.  
 XX  
 PD 12-OCT-2000.  
 XX  
 PF 05-APR-2000; 2000WO-CA000356.  
 XX  
 PR 05-APR-1999; 99US-0127888P.  
 XX  
 PA (UUVI-) UNIV VICTORIA.  
 XX  
 PI White AP, Doran JL, Collison SK, Kay WW;  
 XX  
 DR WPI; 2000-672631/65.  
 DR N-PSDB; AAC64626.  
 XX  
 PT Recombinant agfa gene having a segment replaced by a foreign DNA sequence  
 PT which encodes foreign epitope or antigen, expresses recombinant Agfa  
 PT protein useful for eliciting immune response in animal.  
 XX  
 PS Disclosure; Page 137; 139pp; English.  
 XX  
 CC The present invention describes a recombinant agfa gene (I) where a  
 CC segment of the gene has been replaced by a segment of a foreign DNA  
 CC sequence which encodes a foreign epitope or antigen. Also described are:  
 CC (1) use of thin aggregative fimbriae (SEF17/TAf) nucleation depended  
 CC assembly system of strains of Salmonella, Escherichia coli and  
 CC Enterobacteriaceae for the production of fimbriae comprising recombinant  
 CC Agfa, CsgA and Agfa-homologue fimbria subunits, respectively; (2)  
 CC directing recombination of a recombinant gene into the chromosome of the  
 CC homologous species; (3) directing recombination of a recombinant gene  
 CC back into the chromosome of the homologous species, replacing the native  
 CC copy of that gene; and (4) eliciting an immune response in an animal,  
 CC comprising separating an amino acid polymer comprising a recombinant Agfa  
 CC protein containing a replacement segment or segments of foreign amino  
 CC acid sequence or sequences grown on a Salmonella, E. coli or  
 CC polymer into the animal in conjunction with a carrier or diluent. (I) is  
 CC useful for the expression of recombinant Agfa protein which is useful for  
 CC eliciting an immune response in an animal. In a fimbrial presentation  
 CC system the heterologous antigens are presented in high numbers (up to  
 CC 500,000 copies/cell), the hybrid fimbria protein possesses both the  
 CC immunogenicity and adhesion properties relevant for an efficient live  
 CC vaccine, the carrier fimbrial subunit proteins are usually strong  
 CC immunogens, which may be important for directing an immune response  
 CC against the inserted epitope, and hybrid fimbriae are easy and  
 CC inexpensive to purify in large amount. The present sequence is given in  
 CC the exemplification of the present invention.  
 XX  
 SQ Sequence 151 AA;  
 Query Match 78.4%; Score 611; DB 3; Length 151;  
 Best Local Similarity 80.8%; Pred. No. 2.3e-50;  
 Matches 122; Conservative 7; Mismatches 22; Indels 0; Gaps 0;  
 OY 1 MLLKVAFAAIVVSGSALGAVPQWGGGNNHNGGNSGPDSTLSIYQGSANALALQ 60  
 DB 1 MLLKVAFAAIVVSGSALGAVPQWGGGNNHNGGNSGPDSTLSIYQGSANALALQ 60

```
CC the exemplification of the present invention
XX
SQ Sequence 151 AA;

Query Match 77.7%; Score 605; DB 3; Length 151;
Best Local Similarity 80.1%; Pred. No. 8.5e-50;
Matches 121; Conservative 5; Mismatches 25; Indels 0; Gaps 0;

QY 1 M K L L K V A A F A A I V V G S A L A G V P Q W G G G N H N G G N S S G P D S T L S I Y Q Y G S A N A A L A L Q 60
D b 1 M K L L K V A A F A A I V V G S A L A G V P Q W G G G N H N G G N S S G P D S T L S I Y Q Y G S A N A A L A L Q 60
QY 61 S D A R K S E T T I T Q S G Y G N G A D V G G A D N S T I E L T Q N G F R N N A T I D Q W N A K N S D I T V G Q Y G G 120
D b 61 S D A R K S E T T I T Q S G Y G N G A D V G G A D N S T I E L T Q N G F R N N A T I D Q W N A K N S D I T V G Q Y G G 120
QY 121 N N A A L V N Y D Q L V T R V V T H E M A H A N N A T A N O Y 151
D b 121 N N A A L V N O T A S D S S V M V R Q V G F G N N A T A N O Y 151

RESULT 11
AAB36351
ID AAB36351 standard; protein; 151 AA.
AC AAB36351;
XX
XX 26-FEB-2001 (first entry)
DE Agfa::PT3#6 amino acid sequence SEQ ID NO:22.
XX
XX Salmonella; agfa; chromosomal gene replacement; fimbrin; epitope;
XX vaccine; immune response; immunogen.
XX Salmonella enteritidis.
XX Escherichia coli.
XX Synthetic.
XX WO2000060102-A2.
XX
XX 12-OCT-2000.
XX
XX 05-APR-2000; 2000WO-CA000356.
XX
XX 05-APR-1999; 99US-0127888P.
XX
XX (UYVI-) UNIV VICTORIA.
XX
XX White AP, Doran JL, Collison SK, Kay WW;
XX WPI; 2000-672631/65.
XX N-PSDB; AAC64627.
XX
XX Recombinant agfa gene having a segment replaced by a foreign DNA sequence
XX which encodes foreign epitope or antigen, expresses recombinant Agfa
XX protein useful for eliciting immune response in animal.
XX
XX Disclosure; Page 137; 139pp; English.
XX
XX The present invention describes a recombinant agfa gene (I) where a
XX segment of the gene has been replaced by a segment of a foreign DNA
XX sequence which encodes a foreign epitope or antigen. Also described are:
XX (1) use of thin aggregative fimbriae (SEF17/TAf) nucleation depended
XX assembly system of strains of Salmonella, Escherichia coli and
XX Enterobacteriaceae for the production of fimbriae comprising recombinant
XX Agfa, CsgA and Agfa-homologue fimbrin subunits, respectively; (2)
XX directing recombination of a recombinant gene into the chromosome of the
XX homologous species; (3) directing recombination of a recombinant gene
XX back into the chromosome of the homologous species, replacing the native
XX copy of that gene; and (4) eliciting an immune response in an animal,
XX comprising separating an amino acid polymer comprising a recombinant Agfa
XX protein containing a replacement segment or segments of foreign amino
XX acid sequence or sequences grown on a Salmonella, E. coli or
```

```
CC Enterobacteriaceae host cell, from the host cell and introducing the
CC polymer into the animal in conjunction with a carrier or diluent. (I) is
CC useful for the expression of recombinant Agfa protein which is useful for
CC eliciting an immune response in an animal. In a fimbrial presentation
CC system the heterologous antigens are presented in high numbers (up to
CC 500,000 copies/cell), the hybrid fimbrin protein possesses both the
CC immunogenicity and adhesion properties relevant for an efficient live
CC vaccine, the carrier fimbrial subunit proteins are usually strong
CC immunogens, which may be important for directing an immune response
CC against the inserted epitope, and hybrid fimbriae are easy and
CC inexpensive to purify in large amount. The present sequence is given in
CC the exemplification of the present invention
XX
XX Sequence 151 AA;

Query Match 77.5%; Score 604; DB 3; Length 151;
Best Local Similarity 80.8%; Pred. No. 1.1e-49;
Matches 122; Conservative 6; Mismatches 23; Indels 0; Gaps 0;

QY 1 M K L L K V A A F A A I V V G S A L A G V P Q W G G G N H N G G N S S G P D S T L S I Y Q Y G S A N A A L A L Q 60
D b 1 M K L L K V A A F A A I V V G S A L A G V P Q W G G G N H N G G N S S G P D S T L S I Y Q Y G S A N A A L A L Q 60
QY 61 S D A R K S E T T I T Q S G Y G N G A D V G G A D N S T I E L T Q N G F R N N A T I D Q W N A K N S D I T V G Q Y G G 120
D b 61 S D A R K S E T T I T Q S G Y G N G A D V G G A D N S T I E L T Q N G F R N N A T I D Q W N A K N S D I T V G Q Y G G 120
QY 121 N N A A L V N Y D Q L V T R V V T H E M A H A N N A T A N O Y 151
D b 121 N N A A L V N O T A S D S S V M V R Q V G F G N N A T A N O Y 151

RESULT 12
AAB36355
ID AAB36355 standard; protein; 151 AA.
AC AAB36355;
XX
XX 26-FEB-2001 (first entry)
DE Agfa::PT3#10 amino acid sequence SEQ ID NO:30.
XX
XX Salmonella; agfa; chromosomal gene replacement; fimbrin; epitope;
XX vaccine; immune response; immunogen.
XX Salmonella enteritidis.
XX Escherichia coli.
XX Synthetic.
XX WO2000060102-A2.
XX
XX 12-OCT-2000.
XX
XX 05-APR-2000; 2000WO-CA000356.
XX
XX 05-APR-1999; 99US-0127888P.
XX
XX (UYVI-) UNIV VICTORIA.
XX
XX White AP, Doran JL, Collison SK, Kay WW;
XX WPI; 2000-672631/65.
XX N-PSDB; AAC64631.
XX
XX Recombinant agfa gene having a segment replaced by a foreign DNA sequence
XX which encodes foreign epitope or antigen, expresses recombinant Agfa
XX protein useful for eliciting immune response in animal.
XX
XX Disclosure; Page 139; 139pp; English.
XX
XX The present invention describes a recombinant agfa gene (I) where a
XX segment of the gene has been replaced by a segment of a foreign DNA
XX sequence which encodes a foreign epitope or antigen. Also described are:
```

CC (1) use of thin aggregative fimbriae (SEF17/TAF) nucleation depended  
 CC assembly system of strains of Salmonella, Escherichia coli and  
 CC Enterobacteriaceae for the production of fimbriae comprising recombinant  
 CC Agfa, CsgA and Agfa-homologue fimbrial subunits, respectively; (2)  
 CC directing recombination of a recombinant gene into the chromosome of the  
 CC homologous species; (3) directing recombination of a recombinant gene  
 CC back into the chromosome of the homologous species, replacing the native  
 CC copy of that gene; and (4) eliciting an immune response in an animal,  
 CC comprising separating an amino acid polymer comprising a recombinant Agfa  
 CC protein containing a replacement segment or segments of foreign amino  
 CC acid sequence or sequences grown on a Salmonella, E. coli or  
 CC polymer into the animal in conjunction with a carrier or diluent. (1) is  
 CC useful for the expression of recombinant Agfa protein which is useful for  
 CC eliciting an immune response in an animal. In a fimbrial presentation  
 CC system the heterologous antigens are presented in high numbers (up to  
 CC 500,000 copies/cell), the hybrid fimbrial protein possesses both the  
 CC immunogenicity and adhesion properties relevant for an efficient live  
 CC vaccine, the carrier fimbrial subunit proteins are usually strong  
 CC immunogens, which may be important for directing an immune response  
 CC against the inserted epitope, and hybrid fimbriae are easy and  
 CC inexpensive to purify in large amount. The present sequence is given in  
 CC the exemplification of the present invention

XX Sequence 151 AA;

Query Match 77.4%; Score 603; DB 3; Length 151;  
 Best Local Similarity 80.8%; Pred. No. 1.3e-49;  
 Matches 122; Conservative 6; Mismatches 23; Indels 0; Gaps 0;  
 QY 1 MKLLKVAFAAIVVSGSALAGVVPQWGGGNGHNGSGPDSSTLSIYQYGSANAALALQ 60  
 DB 1 MKLLKVAFAAIVVSGSALAGVVPQWGGGNGHNGSGPDSSTLSIYQYGSANAALALQ 60  
 QY 61 SDARKSETTITQSGYNGADVGQADNSTIELTQNGFRNNATIDQNAKNSDITVGYGG 120  
 DB 61 SDARKSETTITQSGYNGADVGQADNSTIELTQNGFRNNATIDQNAKNSDITVGYGG 120  
 QY 121 NNAALVNYDQLVTRVVTHEMAHANNATANOY 151  
 DB 121 NNAALVNYDQLVTRVVTHEMAHANNATANOY 151

RESULT 13

AAB36348  
 ID AAB36348 standard; protein; 151 AA.

XX AAB36348;  
 XX 26-FEB-2001 (first entry)

DT Agfa::PT3#3 amino acid sequence SEQ ID NO:16.

DE Salmonella; agfa; chromosomal gene replacement; fimbrial; epitope;  
 XX vaccine; immune response; immunogen.

XX Salmonella enteritidis.  
 OS Escherichia coli.  
 OS Synthetic.

XX WO200060102-A2.

XX 12-OCT-2000.

XX 05-APR-2000; 2000WO-CA000356.

XX 05-APR-1999; 99US-0127888P.

XX (UVVI-) UNIV VICTORIA.

XX White AP, Doran JL, Collison SK, Kay WW;

XX WPI; 2000-672631/65.

DR N-PSDB; AAC64624.  
 XX Recombinant agfa gene having a segment replaced by a foreign DNA sequence  
 PT which encodes foreign epitope or antigen, expresses recombinant Agfa  
 PT protein useful for eliciting immune response in animal.

XX Disclosure; Page 136; 139pp; English.

XX The present invention describes a recombinant agfa gene (I) where a  
 CC segment of the gene has been replaced by a segment of a foreign DNA  
 CC sequence which encodes a foreign epitope or antigen. Also described are:  
 CC (1) use of thin aggregative fimbriae (SEF17/TAF) nucleation depended  
 CC assembly system of strains of Salmonella, Escherichia coli and  
 CC Enterobacteriaceae for the production of fimbriae comprising recombinant  
 CC Agfa, CsgA and Agfa-homologue fimbrial subunits, respectively; (2)  
 CC directing recombination of a recombinant gene into the chromosome of the  
 CC homologous species; (3) directing recombination of a recombinant gene  
 CC back into the chromosome of the homologous species, replacing the native  
 CC copy of that gene; and (4) eliciting an immune response in an animal,  
 CC comprising separating an amino acid polymer comprising a recombinant Agfa  
 CC protein containing a replacement segment or segments of foreign amino  
 CC acid sequence or sequences grown on a Salmonella, E. coli or  
 CC polymer into the animal in conjunction with a carrier or diluent. (1) is  
 CC useful for the expression of recombinant Agfa protein which is useful for  
 CC eliciting an immune response in an animal. In a fimbrial presentation  
 CC system the heterologous antigens are presented in high numbers (up to  
 CC 500,000 copies/cell), the hybrid fimbrial protein possesses both the  
 CC immunogenicity and adhesion properties relevant for an efficient live  
 CC vaccine, the carrier fimbrial subunit proteins are usually strong  
 CC immunogens, which may be important for directing an immune response  
 CC against the inserted epitope, and hybrid fimbriae are easy and  
 CC inexpensive to purify in large amount. The present sequence is given in  
 CC the exemplification of the present invention

XX Sequence 151 AA;

Query Match 74.5%; Score 580; DB 3; Length 151;  
 Best Local Similarity 80.1%; Pred. No. 2.1e-47;  
 Matches 121; Conservative 7; Mismatches 23; Indels 0; Gaps 0;

QY 1 MKLLKVAFAAIVVSGSALAGVVPQWGGGNGHNGSGPDSSTLSIYQYGSANAALALQ 60  
 DB 1 MKLLKVAFAAIVVSGSALAGVVPQWGGGNGHNGSGPDSSTLSIYQYGSANAALALQ 60  
 QY 61 SDARKSETTITQSGYNGADVGQADNSTIELTQNGFRNNATIDQNAKNSDITVGYGG 120  
 DB 61 SDARKSETTITQSGYNGADVGQADNSTIELTQNGFRNNATIDQNAKNSDITVGYGG 120  
 QY 121 NNAALVNYDQLVTRVVTHEMAHANNATANOY 151  
 DB 121 NNAALVNYDQLVTRVVTHEMAHANNATANOY 151

RESULT 14

AAR62761  
 ID AAR62761 standard; protein; 120 AA.

XX AAR62761;  
 XX 27-AUG-2003 (revised)  
 DT 25-MAR-2003 (revised)  
 DT 26-JUN-1995 (first entry)

XX Agfa sequence.

XX Salmonella; Agfa; vaccine.

XX Salmonella enteritidis.

XX WO9425598-A2.

XX 10-NOV-1994.

XX 26-APR-1994; 94WO-IB000207.  
PF XX  
XX 26-APR-1993; 93US-00054452.  
PR XX  
XX (UUVI-) UNIV VICTORIA INNOVATION & DEV CORP.  
PA (KING/) KING J.  
PA XX  
XX Kay WW, Collinson SK, Clouthier SC, Doran JL;  
PI XX  
XX WPI; 1994-358275/44.  
DR N-PSDB; AAT73066.  
DR XX  
XX Eliciting an immune response to Salmonella - using attenuated Salmonella  
PT strains, vector constructs, or compenss. contg. fimbrial type proteins.  
PT XX  
PS Disclosure; Fig 7A; 95pp; English.  
PS XX  
XX The sequence represents the Salmonella enteritis 27655-3b TnpH $\alpha$  mutant  
CC strain AgfA protein. The encoding DNA and isolated AgfA protein are used  
CC in genetic immunization and vaccine compositions, respectively, to elicit  
CC an immune response to Salmonella in animals (e.g. food producing animals)  
CC and humans. (Updated on 25-MAR-2003 to correct PN field.) (Updated on 27-  
CC AUG-2003 to correct OS field.)  
XX XX  
XX Sequence 120 AA;  
SQ

Query Match 71.9%; Score 560; DB 2; Length 120;  
Best Local Similarity 100.0%; Pred. No. 1.2e-45;  
Matches 106; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 22 VVPQWGGGNGHNGGSSGPDSTLSIYQYGSANAALALQSDARKSETTITQSGYNGADV 81  
DB 1 VVPQWGGGNGHNGGSSGPDSTLSIYQYGSANAALALQSDARKSETTITQSGYNGADV 60  
QY 82 GQGANSTIELTQNGFRNNAITDQWNAKNSDITVGYGNNALVN 127  
DB 61 GQGANSTIELTQNGFRNNAITDQWNAKNSDITVGYGNNALVN 106

RESULT 15  
AAW23569  
ID AAW23569 standard; protein; 120 AA.  
XX AC  
XX AAW23569;  
XX XX  
XX 25-MAR-2003 (revised)  
DT 29-SEP-1997 (first entry)  
DT XX  
XX Salmonella enteritis 27655-3b TnpH $\alpha$  mutant agfA fragment.  
DE XX  
XX Enteropathogenic bacteria; enterobacteria; S. enteritis; antibody.  
XX XX  
XX Salmonella enteritis.  
XX XX  
XX US5635617-A.  
XX PN  
XX 03-JUN-1997.  
XX XX  
XX 26-APR-1994; 94US-00233788.  
XX PF  
XX 26-APR-1993; 93US-00054452.  
XX PR  
XX (UUVI-) UNIV VICTORIA INNOVATION & DEV CORP.  
PA XX  
XX Collinson SK, Kay WW, Doran JL;  
PI XX  
XX WPI; 1997-309886/28.  
DR N-PSDB; AAT74141.  
DR XX  
XX Isolated Salmonella gene agfA - used for diagnosis of Salmonella or  
PT enteropathogenic bacteria of the Enterobacteria family.  
XX XX

PS Example 2; Fig 7; 85pp; English.  
XX  
XX The present sequence represents an agfA fragment encoded by an agfA gene  
CC fragment derived from Salmonella enteritidis 27655-3b TnpH $\alpha$  mutant  
CC strain. The nucleic acid can be used to provide diagnostic assays for  
CC Salmonella and/or enteropathogenic bacteria of the family Enterobacteria.  
CC It can also be used to provide proteins and antibodies which can be used  
CC for assays. The nucleic acid sequence can be used to provide probes or  
CC primers which can specifically hybridize to nucleic acid molecules from  
CC greater than 99% of Salmonella strains that are pathogenic to warm-  
CC blooded animals relative to nucleic acid molecules from virtually all  
CC other microbial organisms. (Updated on 25-MAR-2003 to correct PF field.)  
XX XX  
XX Sequence 120 AA;  
SQ

Query Match 71.9%; Score 560; DB 2; Length 120;  
Best Local Similarity 100.0%; Pred. No. 1.2e-45;  
Matches 106; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 22 VVPQWGGGNGHNGGSSGPDSTLSIYQYGSANAALALQSDARKSETTITQSGYNGADV 81  
DB 1 VVPQWGGGNGHNGGSSGPDSTLSIYQYGSANAALALQSDARKSETTITQSGYNGADV 60  
QY 82 GQGANSTIELTQNGFRNNAITDQWNAKNSDITVGYGNNALVN 127  
DB 61 GQGANSTIELTQNGFRNNAITDQWNAKNSDITVGYGNNALVN 106

Search completed: August 2, 2004, 14:48:24  
Job time : 45.9 secs





GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: August 2, 2004, 14:40:48 ; Search time 12 Seconds  
(without alignments)  
649.627 Million cell updates/sec

Title: US-09-543-407-12  
Perfect score: 779  
Sequence: 1 MKLLKVAAPAAIVVSGSALA.....VTRVVTHEMAHANNATANYQ 151

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 389414 seqs, 51625971 residues

Total number of hits satisfying chosen parameters: 389414

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Issued Patents AA:\*  
1: /cgn2\_6/ptodata/2/iaa/5A COMB.pdp.\*  
2: /cgn2\_6/ptodata/2/iaa/5B COMB.pdp.\*  
3: /cgn2\_6/ptodata/2/iaa/6A COMB.pdp.\*  
4: /cgn2\_6/ptodata/2/iaa/6B COMB.pdp.\*  
5: /cgn2\_6/ptodata/2/iaa/PTUS COMB.pdp.\*  
6: /cgn2\_6/ptodata/2/iaa/backfiles1.pdp.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	690	88.6	151	1	US-08-233-788A-59
2	560	71.9	120	1	US-08-233-788A-57
3	92	11.8	673	3	US-09-196-387-8
4	92	11.8	673	4	US-09-841-835-8
5	92	11.8	949	3	US-09-196-387-10
6	92	11.8	949	4	US-09-841-835-10
7	92	11.8	1327	3	US-09-196-387-2
8	92	11.8	1327	4	US-09-841-835-2
9	92	11.8	1327	4	US-09-972-115A-8
10	90.5	11.6	745	4	US-09-336-115C-6
11	89.5	11.5	738	3	US-08-864-038A-3
12	89	11.4	943	4	US-09-056-556-204
13	89	11.4	943	4	US-09-072-596-199
14	89	11.4	943	4	US-09-477-135A-131
15	89	11.4	943	4	US-09-072-967-204
16	87.5	11.2	892	4	US-09-336-447A-5
17	85	10.9	956	4	US-09-134-078-63
18	82.5	10.6	186	4	US-09-382-276-3
19	82.5	10.6	873	4	US-09-336-447A-13
20	81.5	10.5	339	4	US-09-252-991A-32096
21	81	10.4	568	4	US-09-543-681A-6966
22	81	10.4	1739	4	US-09-540-236-3739
23	81	10.4	1864	2	US-08-804-227C-3
24	80	10.3	232	4	US-09-252-991A-30263
25	80	10.3	975	4	US-09-328-352-4764
26	80	10.3	3241	4	US-09-841-786-1
27	79.5	10.2	361	4	US-09-540-236-2164

28	79	10.1	941	4	US-09-336-447A-9	Sequence 9, Appli
29	78.5	10.1	2123	3	US-08-968-685A-10	Sequence 10, Appli
30	78	10.0	906	1	US-08-254-573-2	Sequence 2, Appli
31	78	10.0	906	1	US-08-687-379-2	Sequence 4, Appli
32	78	10.0	906	1	US-08-687-379-4	Sequence 1, Appli
33	78	10.0	906	1	US-08-172-332-1	Sequence 2, Appli
34	78	10.0	906	1	US-08-216-326-2	Sequence 22119, A
35	77.5	9.9	702	4	US-09-252-991A-22119	Sequence 4, Appli
36	77.5	9.9	714	4	US-09-841-786-4	Sequence 3401, Ap
37	77.5	9.9	878	4	US-09-540-236-3401	Sequence 280, App
38	77	9.9	415	4	US-09-025-769B-280	Sequence 8, Appli
39	77	9.9	528	4	US-09-490-291-8	Sequence 39, Appli
40	77	9.9	1690	4	US-09-595-684B-39	Sequence 5434, Ap
41	77	9.9	2315	4	US-09-543-681A-5434	Sequence 2, Appli
42	76.5	9.8	159	3	US-08-856-253-2	Sequence 2, Appli
43	76.5	9.8	186	4	US-09-382-276-2	Sequence 13002, A
44	76.5	9.8	605	4	US-09-489-039A-13002	Sequence 7, Appli
45	75.5	9.7	624	4	US-09-336-447A-7	

ALIGNMENTS

RESULT 1  
US-08-233-788A-59  
; Sequence 59, Application US/08233788A  
; Patent No. 5635617  
; GENERAL INFORMATION:  
; APPLICANT: Doran, James L.  
; APPLICANT: Kay, William W.  
; APPLICANT: Collinson, Sharon S.  
; APPLICANT: Clouthier, Sharon S.  
; TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR DETECTION  
; TITLE OF INVENTION: OF SALMONELLA  
; NUMBER OF SEQUENCES: 61  
; CORRESPONDENCE ADDRESS:  
; ADDRESSES: Seed and Berry  
; STREET: 6300 Columbia Center, 701 Fifth Avenue  
; CITY: Seattle  
; STATE: Washington  
; COUNTRY: U.S.A.  
; ZIP: 98104-7092  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/233,788A  
; FILING DATE: 26-APR-1994  
; CLASSIFICATION: 435  
; ATTORNEY/AGENT INFORMATION:  
; NAME: King, Joshua  
; REGISTRATION NUMBER: 35,570  
; REFERENCE/DOCKET NUMBER: 920043.403C2  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (206) 682-6031  
; TELEFAX: (206) 622-4900  
; TELEX: 3723836 SEEDANBERRY  
; INFORMATION FOR SEQ ID NO: 59:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 151 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
; US-08-233-788A-59

Query Match 88.6%; Score 690; DB 1; Length 151;  
Best Local Similarity 89.4%; Pred. No. 7, 1e-62;  
Matches 135; Conservative 12; Indels 0; Gaps 0;

QY 1 MKLLKVAAPAAIVVSGSALAGVTPWGGGHHNGGSGPDPSTLSIYVYGSANALALQ 60

```

Db 1 MLLKVAFAAIIVVSGSALAGVVPQGGGNGHNGSGPDSLTLSIYQGSANAALALQ 60
Qy 61 SDARKSETTITQSGYNGADVGQADNSTIELTQNGFRNNATIDQNAKNSDITVQYGG 120
Db 61 SDARKSETTITQSGYNGADVGQADNSTIELTQNGFRNNATIDQNAKNSDITVQYGG 120
Qy 121 NNALVNYDQLVTRVTHMAHANNATQY 151
Db 121 NNPALVNTQADSSVMVRQVGFNNATQY 151

RESULT 2
US-08-233-788A-57
; Sequence 57, Application US/08233788A
; Patent No. 5635617
; GENERAL INFORMATION:
; APPLICANT: Doran, James L.
; APPLICANT: Kay, William W.
; APPLICANT: Collinson, Karen S.
; APPLICANT: Clouthier, Sharon C.
; TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR DETECTION
; TITLE OF INVENTION: OF SALMONELLA
; NUMBER OF SEQUENCES: 61
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Seed and Berry
; STREET: 6300 Columbia Center, 701 Fifth Avenue
; CITY: Seattle
; STATE: Washington
; COUNTRY: U.S.A.
; ZIP: 98104-7092
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/233,788A
; FILING DATE: 26-APR-1994
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: King, Joshua
; REGISTRATION NUMBER: 35,570
; REFERENCE/DOCKET NUMBER: 920043.403C2
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (206) 622-4900
; TELEFAX: (206) 682-6031
; TELEX: 3723836 SEEDANBERRY
; INFORMATION FOR SEQ ID NO: 57:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 120 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-233-788A-57

Query Match 71.9%; Score 560; DB 1; Length 120;
Best Local Similarity 100.0%; Pred. No. 6e-49;
Matches 106; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 22 VVPQGGGNGHNGSGPDSLTLSIYQGSANAALALQSDARKSETTITQSGYNGADV 81
Db 1 VVPQGGGNGHNGSGPDSLTLSIYQGSANAALALQSDARKSETTITQSGYNGADV 60

Qy 82 GQADNSTIELTQNGFRNNATIDQNAKNSDITVQYGGNNAALVN 127
Db 61 GQADNSTIELTQNGFRNNATIDQNAKNSDITVQYGGNNAALVN 106

RESULT 3
US-09-196-387-8
; Sequence 8, Application US/09196387
; Patent No. 6277613
; GENERAL INFORMATION:

```

```

; APPLICANT: de Lange, Titia
; APPLICANT: Smith, Susan
; TITLE OF INVENTION: A PROTEIN THAT BINDS TO TRF1 AND METHODS
; TITLE OF INVENTION: OF USE THEREOF
; NUMBER OF SEQUENCES: 12
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Klauber & Jackson
; STREET: 411 Hackensack Avenue, 4th Floor
; CITY: Hackensack
; STATE: New Jersey
; COUNTRY: USA
; ZIP: 07601
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/196,387
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 09/095,225
; FILING DATE: June 10, 1998
; ATTORNEY/AGENT INFORMATION:
; NAME: Jackson Esq., David A.
; REGISTRATION NUMBER: 26,742
; REFERENCE/DOCKET NUMBER: 600-1-230 CIP1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 201-487-5800
; TELEFAX: 201-343-1684
; TELEX: 133521
; INFORMATION FOR SEQ ID NO: 8:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 673 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-09-196-387-8

Query Match 11.8%; Score 92; DB 3; Length 673;
Best Local Similarity 30.4%; Pred. No. 0.6;
Matches 35; Conservative 15; Mismatches 57; Indels 8; Gaps 4;

Qy 6 VAAFAAI-VVSGSALAGVVPQGGGNGHNGSGPDSLTLSIYQGSANAALALQSDAR 64
Db 99 VAAAPVVPVAVSTSSAAGVAPNPAGSGNNPSSSPSSPTSS-SSSSPSSPGSLAESPEAA 157

Qy 65 KSETTIT---QSGYNGADVQCADNSTIELTQNG--FRNNATIDQNAKNSDI 113
Db 158 GVSSTAPLPGGAAGPGTGVPVAVSGALRELLACRNGDVSRVKRLVDAANYNAKDM 212

RESULT 4
US-09-841-835-8
; Sequence 8, Application US/09841835
; Patent No. 6506587
; GENERAL INFORMATION:
; APPLICANT: de Lange, Titia
; APPLICANT: Smith, Susan
; TITLE OF INVENTION: A PROTEIN THAT BINDS TO TRF1 AND METHODS
; TITLE OF INVENTION: OF USE THEREOF
; NUMBER OF SEQUENCES: 12
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Klauber & Jackson
; STREET: 411 Hackensack Avenue, 4th Floor
; CITY: Hackensack
; STATE: New Jersey
; COUNTRY: USA
; ZIP: 07601
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible

```

OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/841,835  
FILING DATE:  
CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 09/196,387  
FILING DATE:  
ATTORNEY/AGENT INFORMATION:  
NAME: Jackson Esq., David A.  
REGISTRATION NUMBER: 26,742  
REFERENCE/DOCKET NUMBER: 600-1-230 CIP1  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 201-487-5800  
TELEFAX: 201-343-1684  
INFORMATION FOR SEQ ID NO: 8:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 673 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-09-841-835-8

Query Match 11.8%; Score 92; DB 4; Length 673;  
Best Local Similarity 30.4%; Pred. No. 0.6; Mismatches 15; Indels 8; Gaps 4;  
Matches 35; Conservative 15;  
QY 6 VAAFAAI-VVSGSALAGVVPQWGGGNGGNNHGGSGPDSTLSIYQGSANAALALQSDAR 64  
DB 99 VAAAPVPAVSTSSAAGVAPNAPAGSGNNGSPSSSPTSS-SSSSPSPGSLAESPEAA 157  
QY 65 KSETTIT-----QSGYNGADVQGGADNSTIELTQNG--FRNNATIDOWNAKNSDI 113  
DB 158 GVSSTAPLPGAGPGTGVPAVSGALRELLACRNGDYSRVKRLVDAANVNAKDM 212

RESULT 5  
US-09-196-387-10  
Sequence 10, Application US/09196387  
Patent No. 6277613  
GENERAL INFORMATION:  
APPLICANT: de Lange, Titia  
APPLICANT: Smith, Susan  
TITLE OF INVENTION: A PROTEIN THAT BINDS TO TRF1 AND METHODS  
NUMBER OF SEQUENCES: 12  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Klauber & Jackson  
STREET: 411 Hackensack Avenue, 4th Floor  
CITY: Hackensack  
STATE: New Jersey  
COUNTRY: USA  
ZIP: 07601  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/196,387  
FILING DATE:  
CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 09/095,225  
FILING DATE: June 10, 1998  
ATTORNEY/AGENT INFORMATION:  
NAME: Jackson Esq., David A.  
REGISTRATION NUMBER: 26,742  
REFERENCE/DOCKET NUMBER: 600-1-230 CIP1  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 201-487-5800

TELEFAX: 201-343-1684  
TELEX: 133521  
INFORMATION FOR SEQ ID NO: 10:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 949 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-09-196-387-10

Query Match 11.8%; Score 92; DB 3; Length 949;  
Best Local Similarity 30.4%; Pred. No. 0.94;  
Matches 35; Conservative 15; Mismatches 57; Indels 8; Gaps 4;  
QY 6 VAAFAAI-VVSGSALAGVVPQWGGGNGGNNHGGSGPDSTLSIYQGSANAALALQSDAR 64  
DB 99 VAAAPVPAVSTSSAAGVAPNAPAGSGNNGSPSSSPTSS-SSSSPSPGSLAESPEAA 157  
QY 65 KSETTIT-----QSGYNGADVQGGADNSTIELTQNG--FRNNATIDOWNAKNSDI 113  
DB 158 GVSSTAPLPGAGPGTGVPAVSGALRELLACRNGDYSRVKRLVDAANVNAKDM 212

RESULT 6  
US-09-841-835-10  
Sequence 10, Application US/09841835  
Patent No. 6506587  
GENERAL INFORMATION:  
APPLICANT: de Lange, Titia  
APPLICANT: Smith, Susan  
TITLE OF INVENTION: A PROTEIN THAT BINDS TO TRF1 AND METHODS  
NUMBER OF SEQUENCES: 12  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Klauber & Jackson  
STREET: 411 Hackensack Avenue, 4th Floor  
CITY: Hackensack  
STATE: New Jersey  
COUNTRY: USA  
ZIP: 07601  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/841,835  
FILING DATE:  
CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 09/196,387  
FILING DATE:  
ATTORNEY/AGENT INFORMATION:  
NAME: Jackson Esq., David A.  
REGISTRATION NUMBER: 26,742  
REFERENCE/DOCKET NUMBER: 600-1-230 CIP1  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 201-487-5800  
TELEFAX: 201-343-1684  
TELEX: 133521  
INFORMATION FOR SEQ ID NO: 10:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 949 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-09-841-835-10

Query Match 11.8%; Score 92; DB 4; Length 949;  
Best Local Similarity 30.4%; Pred. No. 0.94;  
Matches 35; Conservative 15; Mismatches 57; Indels 8; Gaps 4;  
QY 6 VAAFAAI-VVSGSALAGVVPQWGGGNGGNNHGGSGPDSTLSIYQGSANAALALQSDAR 64

Db 99 VAAAPVPAVSTSSAAGVAPNAGSGNNPSSSSPTSS-SSSSPSPGSSLAESPEAA 157  
Qy 65 KSETTIT-----QSGYGNAGVGGAGDNSTIELTQNG--FRNNATIDQWNAKNSDI 113  
Db 158 GVSSTAPLPGCAAGPGTGVPVAVSGALRELLACRNGDVSVRKVLVDAAVNAKDM 212

## RESULT 7

US-09-196-387-2  
; Sequence 2, Application US/09196387  
; Patent No. 6277613  
; GENERAL INFORMATION:  
; APPLICANT: de Lange, Titia  
; TITLE OF INVENTION: A PROTEIN THAT BINDS TO TRF1 AND METHODS  
; TITLE OF INVENTION: OF USE THEREOF  
; NUMBER OF SEQUENCES: 12  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Klauber & Jackson  
; STREET: 411 Hackensack Avenue, 4th Floor  
; CITY: Hackensack  
; STATE: New Jersey  
; COUNTRY: USA  
; ZIP: 07601  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/09/196,387  
; FILING DATE:  
; CLASSIFICATION:  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: 09/196,387  
; FILING DATE:  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Jackson Esq., David A.  
; REGISTRATION NUMBER: 26,742  
; REFERENCE/DOCKET NUMBER: 600-1-230 CIP1  
; TELEPHONE: 201-487-5800  
; TELEFAX: 201-343-1684  
; TELEX: 133521  
; INFORMATION FOR SEQ ID NO: 2:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 1327 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
; HYPOTHETICAL: NO  
US-09-196-387-2

Query Match 11.8%; Score 92; DB 3; Length 1327;  
Best Local Similarity 30.4%; Pred. No. 1.5;  
Matches 35; Conservative 15; Mismatches 57; Indels 8; Gaps 4;  
Qy 6 VAAFAAI-VVSGSALAGVVPWGGGNGHNGSGSPDSTLSIYQVGSANAALALQSDAR 64  
Db 99 VAAAPVPAVSTSSAAGVAPNAGSGNNPSSSSPTSS-SSSSPSPGSSLAESPEAA 157  
Qy 65 KSETTIT-----QSGYGNAGVGGAGDNSTIELTQNG--FRNNATIDQWNAKNSDI 113  
Db 158 GVSSTAPLPGCAAGPGTGVPVAVSGALRELLACRNGDVSVRKVLVDAAVNAKDM 212

## RESULT 8

US-09-841-835-2  
; Sequence 2, Application US/09841835  
; Patent No. 6506587  
; GENERAL INFORMATION:

; APPLICANT: de Lange, Titia  
; APPLICANT: Smith, Susan  
; TITLE OF INVENTION: A PROTEIN THAT BINDS TO TRF1 AND METHODS  
; TITLE OF INVENTION: OF USE THEREOF  
; NUMBER OF SEQUENCES: 12  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Klauber & Jackson  
; STREET: 411 Hackensack Avenue, 4th Floor  
; CITY: Hackensack  
; STATE: New Jersey  
; COUNTRY: USA  
; ZIP: 07601  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/09/841,835  
; FILING DATE:  
; CLASSIFICATION:  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: 09/196,387  
; FILING DATE:  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Jackson Esq., David A.  
; REGISTRATION NUMBER: 26,742  
; REFERENCE/DOCKET NUMBER: 600-1-230 CIP1  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 201-487-5800  
; TELEFAX: 201-343-1684  
; TELEX: 133521  
; INFORMATION FOR SEQ ID NO: 2:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 1327 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
; HYPOTHETICAL: NO  
US-09-841-835-2

Query Match 11.8%; Score 92; DB 4; Length 1327;  
Best Local Similarity 30.4%; Pred. No. 1.5;  
Matches 35; Conservative 15; Mismatches 57; Indels 8; Gaps 4;  
Qy 6 VAAFAAI-VVSGSALAGVVPWGGGNGHNGSGSPDSTLSIYQVGSANAALALQSDAR 64  
Db 99 VAAAPVPAVSTSSAAGVAPNAGSGNNPSSSSPTSS-SSSSPSPGSSLAESPEAA 157  
Qy 65 KSETTIT-----QSGYGNAGVGGAGDNSTIELTQNG--FRNNATIDQWNAKNSDI 113  
Db 158 GVSSTAPLPGCAAGPGTGVPVAVSGALRELLACRNGDVSVRKVLVDAAVNAKDM 212

## RESULT 9

US-09-972-115A-8  
; Sequence 8, Application US/09972115A  
; Patent No. 6599728  
; GENERAL INFORMATION:  
; APPLICANT: Geron Corporation  
; APPLICANT: Gregg, Morin B.  
; APPLICANT: Walter, Funk D.  
; APPLICANT: Mieczyslaw, Platysek A.  
; TITLE OF INVENTION: A Second Mammalian Telomerase  
; FILE REFERENCE: 080/003C  
; CURRENT APPLICATION NUMBER: US/09/972,115A  
; CURRENT FILING DATE: 2001-10-05  
; PRIOR APPLICATION NUMBER: US 60/128,577  
; PRIOR FILING DATE: 2000-04-10  
; PRIOR APPLICATION NUMBER: US 60/129,123  
; PRIOR FILING DATE: 1999-04-13  
; NUMBER OF SEQ ID NOS: 64

```

; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 8
; LENGTH: 1327
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-972-115A-8

Query Match      11.8%; Score 92; DB 4; Length 1327;
Best Local Similarity 30.4%; Pred. No. 1.5;
Matches 35; Conservative 15; Mismatches 57; Indels 8; Gaps 4;

QY 6 VAAFAAI-VVGSALAGVVPQWGGGNGHNGGNSGPDSTLSIYQVGSANAALQSDAR 64
DB 99 VAAAPVPAVTSAAAGVAPNPAGSGNSNPSSSSPTSS-SSSPSPSPGSLAESPEAA 157
QY 65 KSETTIT-OSGYGNGADVGGADNNTIETQNG--FRNNATIDWNKNSDI 113
DB 158 GVSSTAPLPGGAAGPGTGVPAVSGALRELLACRNGDVSVKRLVDAAVNNAKDM 212

RESULT 10.
US-09-336-115C-6
; Sequence 6, Application US/09336115C
; Patent No. 6576244
; GENERAL INFORMATION:
; APPLICANT: Weltzin, Richard A.
; TITLE OF INVENTION: LT and CT in Parenteral Immunization
; FILE REFERENCE: 06132/055002
; CURRENT APPLICATION NUMBER: US/09/336,115C
; CURRENT FILING DATE: 1999-06-18
; PRIOR APPLICATION NUMBER: US 09/100,258
; PRIOR FILING DATE: 1998-06-19
; NUMBER OF SEQ ID NOS: 24
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 6
; LENGTH: 745
; TYPE: PRT
; ORGANISM: Helicobacter pylori
; FEATURE:
; NAME/KEY: SIGNAL
; LOCATION: (1)...(20)
; FEATURE:
; NAME/KEY: VARIANT
; LOCATION: 721
; OTHER INFORMATION: Xaa = Any Amino Acid
US-09-336-115C-6

Query Match      11.6%; Score 90.5; DB 4; Length 745;
Best Local Similarity 20.8%; Pred. No. 0.97;
Matches 40; Conservative 28; Mismatches 63; Indels 61; Gaps 8;

QY 8 AFAIVVGSALAGVVPQW-----GGGNGHNGGN----- 37
DB 102 AYQAVFLAINAAVGL---WNTIGVAVMCGNGNGTSGFGSVIENDPQGDSTOITCNRFE 158
QY 38 SSGPDSTLSIYQVGSANAALQSDARKSETTITQSGYNG-----ADVQG 83
DB 159 STGCKSMSIDFKKLEAYVLIQKALKNQSGPPELG-GNGTKVSVNMYNVECRQTADING 217
QY 84 G-----ADNSTIELTQNGFRNNATIDWNKNSDIYQVGGNNAALVNYDQLVTVVT 137
DB 218 GVYQFCRAKNGS-SSSSNGGNGSQTATTQDGVITTTTNNKATVKED-----IT 270
QY 138 HEMAANNATAN 149
DB 271 NNAEQLLNQAAAN 282

RESULT 11
US-08-864-038A-3
; Sequence 3, Application US/08864038A
```

```

; Patent No. 6001592
; GENERAL INFORMATION:
; APPLICANT: Kunio NAKASHIMA et al.
; TITLE OF INVENTION: NOVEL POLYPEPTIDE GENE CDNA, VECTOR
; CONTAINING SAID CDNA, HOST CELLS TRANSFORMED WITH SAID
; VECTOR, POLYPEPTIDE PRODUCED THEREBY, METHOD OF PRODUCING
; SAID POLYPEPTIDE, DNA ENCODING SAID POLYPEPTIDE AND ANTIBODY
; TO SAID POLYPEPTIDE
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: 812-5 Hirano
; STREET: Isshinden
; CITY: Tsu-city
; STATE: Mie-prefecture
; COUNTRY: JAPAN
; ZIP: 514-01
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.50 inch, 1.44 MB storage
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: Microsoft Windows 95
; SOFTWARE: Word Perfect 6.1
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/864,038A
; FILING DATE: May 28, 1997
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: JP 8-184459
; FILING DATE: 15-July-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: C. Bruce Hamburg
; REGISTRATION NUMBER: 22,389
; REFERENCE/DOCKET NUMBER: F-5610
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212)986-2340
; TELEFAX: (212)953-7733
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; ORIGINAL SOURCE:
; ORGANISM: Pinctada fucata
; CELL TYPE: mantle epithelial cell
; FEATURE:
; NAME/KEY: peptide
; LOCATION: from 1 to 738
; IDENTIFICATION METHOD: E (by experiment)
; US-08-864-038A-3

Query Match      11.5%; Score 89.5; DB 3; Length 738;
Best Local Similarity 35.4%; Pred. No. 1.2;
Matches 29; Conservative 4; Mismatches 23; Indels 23; Gaps 3;

QY 3 LLKVAFAAIVVGSALAGVVPQWGGGNGHNGGNSGPDSTLSIYQVGSANAALQSD 62
DB 419 LLKSSASASASASASAG-----GGGGGGGGGGGG-----GGGAGALA----- 460
QY 63 ARKSETTITQSGYNGADVQGG 84
DB 461 -----AALAAAGAGGGGLGGGG 477

RESULT 12
US-09-056-556-204
; Sequence 204, Application US/09056556
; Patent No. 6350456
; GENERAL INFORMATION:
; APPLICANT: Reed, Steven G.
; APPLICANT: Skeiky, Yasir A.W.
; APPLICANT: Dillon, Davin C.
; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR THE PREVENTION AND
; NUMBER OF SEQUENCES: 241
; CORRESPONDENCE ADDRESS:
```

ADDRESS: SEED and BERRY LLP  
 STREET: 6300 Columbia Center, 701 Fifth Avenue  
 CITY: Seattle  
 STATE: Washington  
 COUNTRY: USA  
 ZIP: 98104-7092  
 COMPUTER READABLE FORM:  
 MEDIUM TYPE: Floppy disk  
 OPERATING SYSTEM: PC-DOS/MS-DOS  
 SOFTWARE: PatentIn Release #1.0, Version #1.30  
 CURRENT APPLICATION DATA:  
 APPLICATION NUMBER: US/09/056,556  
 FILING DATE: 07-APR-1998  
 CLASSIFICATION:  
 ATTORNEY/AGENT INFORMATION:  
 NAME: Maki, David J.  
 REGISTRATION NUMBER: 31,392  
 REFERENCE/DOCKET NUMBER: 210121.457  
 TELEPHONE: (206) 622-4900  
 TELEFAX: (206) 682-6031  
 INFORMATION FOR SEQ ID NO: 204:  
 SEQUENCE CHARACTERISTICS:  
 LENGTH: 943 amino acids  
 TYPE: amino acid  
 STRANDEDNESS:  
 TOPOLOGY: linear  
 US-09-056-556-204

Query Match 11.4%; Score 89; DB 4; Length 943;  
 Best Local Similarity 25.6%; Pred. No. 1.9;  
 Matches 33; Conservative 15; Mismatches 55; Indels 26; Gaps 5;  
 QY 12 IVVSGSALAGVVPWGGGNGHNGGSSGPDSTLSIYQYGSANAALALQSDAR---KSET 68  
 DB 159 IGLTSGLLGF-----GGLNSGTGN-----IGLFNSGTGNVGNISGTGNWIGNSG 205  
 QY 69 TITQSGYGNAGDVQGGADNSTIETQNGFRNATIDOWNAKNSD---ITVGQY----- 118  
 DB 206 NSYTGFGNSGDANTGFNSGIANTGVGNAGNTYNTGPNNSNTGSGFNWQYNTGYLNS 265  
 QY 119 GGNNAALVN 127  
 DB 266 GNYNTGLAN 274

RESULT 13  
 US-09-072-596-199  
 Sequence 199, Application US/09072596  
 Patent No. 6458366  
 GENERAL INFORMATION:  
 APPLICANT: Reed, Steven G.  
 APPLICANT: Skeiky, Yasir A.W.  
 APPLICANT: Dillon, David C.  
 APPLICANT: Campos-Neto, Antonia  
 APPLICANT: Houghton, Raymond  
 APPLICANT: Vedwick, Thomas S.  
 APPLICANT: Twardzik, Daniel R.  
 APPLICANT: Lodes, Michael J.  
 APPLICANT: Hendrickson, Ronald C.  
 TITLE OF INVENTION: COMPOUNDS AND METHODS FOR DIAGNOSIS OF  
 NUMBER OF SEQUENCES: 350  
 CORRESPONDENCE ADDRESS:  
 ADDRESS: SEED and BERRY LLP  
 STREET: 6300 Columbia Center, 701 Fifth Avenue  
 CITY: Seattle  
 STATE: Washington  
 COUNTRY: USA  
 ZIP: 98104-7092  
 COMPUTER READABLE FORM:  
 MEDIUM TYPE: Floppy disk  
 COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS  
 SOFTWARE: PatentIn Release #1.0, Version #1.30  
 CURRENT APPLICATION DATA:  
 APPLICATION NUMBER: US/09/072,596  
 FILING DATE: 05-MAY-1998  
 CLASSIFICATION:  
 ATTORNEY/AGENT INFORMATION:  
 NAME: Maki, David J.  
 REGISTRATION NUMBER: 31,392  
 REFERENCE/DOCKET NUMBER: 210121.417C9  
 TELECOMMUNICATION INFORMATION:  
 TELEPHONE: (206) 622-4900  
 TELEFAX: (206) 682-6031  
 INFORMATION FOR SEQ ID NO: 199:  
 SEQUENCE CHARACTERISTICS:  
 LENGTH: 943 amino acids  
 TYPE: amino acid  
 STRANDEDNESS:  
 TOPOLOGY: linear  
 US-09-072-596-199

Query Match 11.4%; Score 89; DB 4; Length 943;  
 Best Local Similarity 25.6%; Pred. No. 1.9;  
 Matches 33; Conservative 15; Mismatches 55; Indels 26; Gaps 5;  
 QY 12 IVVSGSALAGVVPWGGGNGHNGGSSGPDSTLSIYQYGSANAALALQSDAR---KSET 68  
 DB 159 IGLTSGLLGF-----GGLNSGTGN-----IGLFNSGTGNVGNISGTGNWIGNSG 205  
 QY 69 TITQSGYGNAGDVQGGADNSTIETQNGFRNATIDOWNAKNSD---ITVGQY----- 118  
 DB 206 NSYTGFGNSGDANTGFNSGIANTGVGNAGNTYNTGPNNSNTGSGFNWQYNTGYLNS 265  
 QY 119 GGNNAALVN 127  
 DB 266 GNYNTGLAN 274

RESULT 14  
 US-09-477-135A-131  
 Sequence 131, Application US/09477135A  
 Patent No. 6572865  
 GENERAL INFORMATION:  
 APPLICANT: Nano, Francis  
 TITLE OF INVENTION: Mycobacterium Tuberculosis DNA Sequences Encoding  
 TITLE OF INVENTION: Immunostimulatory Peptides  
 FILE REFERENCE: 52888  
 CURRENT APPLICATION NUMBER: US/09/477,135A  
 CURRENT FILING DATE: 2000-01-03  
 PRIOR APPLICATION NUMBER: 0890823  
 PRIOR FILING DATE: 1997-12-15  
 PRIOR APPLICATION NUMBER: US 96/10375  
 PRIOR FILING DATE: 1996-06-14  
 PRIOR APPLICATION NUMBER: 60/000,254  
 PRIOR FILING DATE: 1995-06-15  
 NUMBER OF SEQ ID NOS: 169  
 SOFTWARE: PatentIn Ver. 2.0  
 SEQ ID NO 131  
 LENGTH: 943  
 TYPE: PRT  
 ORGANISM: Mycobacterium tuberculosis  
 US-09-477-135A-131

Query Match 11.4%; Score 89; DB 4; Length 943;  
 Best Local Similarity 25.6%; Pred. No. 1.9;  
 Matches 33; Conservative 15; Mismatches 55; Indels 26; Gaps 5;  
 QY 12 IVVSGSALAGVVPWGGGNGHNGGSSGPDSTLSIYQYGSANAALALQSDAR---KSET 68  
 DB 266 IGLTSGLLGF-----GGLNSGTGN-----IGLFNSGTGNVGNISGTGNWIGNSG 312  
 QY 69 TITQSGYGNAGDVQGGADNSTIETQNGFRNATIDOWNAKNSD---ITVGQY----- 118

Job time : 13 secs

Db 313 NSYNTGFGSGDANTGFFNSGIANTGVGNAGNTGSYNPGNSNTGGFNMGGQYNTGYLNS 372

Qy 119 GGNNAALVN 127  
Db 373 GNYNTGLAN 381

RESULT 15  
US-09-072-967-204  
; Sequence 204, Application US/09072967  
; Patent No. 6592877  
; GENERAL INFORMATION:  
; APPLICANT: Reed, Steven G.  
; APPLICANT: Skeiky, Yasir A.W.  
; APPLICANT: Dillon, Davin C.  
; APPLICANT: Campos-Neto, Antonio  
; APPLICANT: Houghton, Raymond  
; APPLICANT: Vedwick, Thomas S.  
; APPLICANT: Twardzik, Daniel R.  
; APPLICANT: Lodes, Michael J.  
; APPLICANT: Hendrickson, Ronald C.  
; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR IMMUNOTHERAPY  
; TITLE OF INVENTION: AND DIAGNOSIS OF TUBERCULOSIS  
; NUMBER OF SEQUENCES: 355  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: SEED and BERRY LLP  
; STREET: 6300 Columbia Center, 701 Fifth Avenue  
; CITY: Seattle  
; STATE: Washington  
; COUNTRY: USA  
; ZIP: 98104-7092  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/09/072.967  
; FILING DATE: 05-MAY-1998  
; CLASSIFICATION:  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Maki, David J.  
; REGISTRATION NUMBER: 31,392  
; REFERENCE/DOCKET NUMBER: 210121.411C9  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (206) 622-4900  
; TELEFAX: (206) 682-6031  
; INFORMATION FOR SEQ ID NO: 204:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 943 amino acids  
; TYPE: amino acid  
; STRANDEDNESS:  
; TOPOLOGY: linear  
US-09-072-967-204

Query Match 11.4%; Score 89; DB 4; Length 943;  
Best Local Similarity 25.6%; Pred.No.1.9;  
Matches 33; Conservative 15; Mismatches 55; Indels 26; Gaps 5;  
Qy 12 IWVSGSALAGVVPQWGGGNGHGGSGSDSTLSIYQYGSANAALALQSDAR---KSET 68  
Db 159 IGLTSGLLGF-----GGLNSGTGN-----IGLFSGTGNYGIGNSGTGNWIGNSG 205  
Qy 69 TITSGYNGADVGGGADNSTIELTONGFRNATIDOWNAKNSD---ITVQY----- 118  
Db 206 NSYNTGFGSGDANTGFFNSGIANTGVGNAGNTGSYNPGNSNTGGFNMGGQYNTGYLNS 265  
Qy 119 GGNNAALVN 127  
Db 266 GNYNTGLAN 274

Search completed: August 2, 2004, 14:58:31

This Page Blank (uspto)



GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: August 2, 2004, 14:54:48 ; Search time 36.8 Seconds  
(without alignments)  
1287.123 Million cell updates/sec

Title: US-09-543-407-12

Perfect score: 779

Sequence: 1 MKLLKVAFAAIVVSSGSLA.....VTRVVTMHHANNATANCY 151

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1291235 seqs, 313682936 residues

Total number of hits satisfying chosen parameters: 1291235

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

Published Applications\_AA:\*  
1: /cgn2\_6/prodata/2/pubpaa/US07\_PUBCOMB.pep.\*  
2: /cgn2\_6/prodata/2/pubpaa/PCT\_NEW\_PUB.pep.\*  
3: /cgn2\_6/prodata/2/pubpaa/US06\_NEW\_PUB.pep.\*  
4: /cgn2\_6/prodata/2/pubpaa/US06\_PUBCOMB.pep.\*  
5: /cgn2\_6/prodata/2/pubpaa/US07\_NEW\_PUB.pep.\*  
6: /cgn2\_6/prodata/2/pubpaa/PCTUS\_PUBCOMB.pep.\*  
7: /cgn2\_6/prodata/2/pubpaa/US08\_NEW\_PUB.pep.\*  
8: /cgn2\_6/prodata/2/pubpaa/US08\_PUBCOMB.pep.\*  
9: /cgn2\_6/prodata/2/pubpaa/US09A\_PUBCOMB.pep.\*  
10: /cgn2\_6/prodata/2/pubpaa/US09B\_PUBCOMB.pep.\*  
11: /cgn2\_6/prodata/2/pubpaa/US09C\_PUBCOMB.pep.\*  
12: /cgn2\_6/prodata/2/pubpaa/US09\_NEW\_PUB.pep.\*  
13: /cgn2\_6/prodata/2/pubpaa/US10A\_PUBCOMB.pep.\*  
14: /cgn2\_6/prodata/2/pubpaa/US10B\_PUBCOMB.pep.\*  
15: /cgn2\_6/prodata/2/pubpaa/US10C\_PUBCOMB.pep.\*  
16: /cgn2\_6/prodata/2/pubpaa/US10\_NEW\_PUB.pep.\*  
17: /cgn2\_6/prodata/2/pubpaa/US60\_NEW\_PUB.pep.\*  
18: /cgn2\_6/prodata/2/pubpaa/US60\_PUBCOMB.pep.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	517	66.4	151	12	US-09-741-873B-4
2	517	66.4	151	12	US-09-741-873B-4
3	439	56.4	131	12	US-09-741-873B-2
4	439	56.4	131	12	US-09-741-873B-2
5	100	12.8	445	15	US-10-369-493-20638
6	97.5	12.5	263	12	US-10-425-114-49960
7	94.5	12.1	678	12	US-10-282-122A-64573
8	94.5	12.1	688	14	US-10-032-585-7876
9	93	11.9	271	16	US-10-437-963-147343
10	92	11.8	673	9	US-09-841-835-8
11	92	11.8	949	9	US-09-841-835-10
12	92	11.8	1327	9	US-09-841-835-2
13	92	11.8	1327	10	US-09-972-115A-8
14	92	11.8	1327	14	US-10-199-937-4
15	91.5	11.7	705	16	US-10-437-963-203823

Sequence 162284, Sequence 6, Appli  
Sequence 58683, A  
Sequence 749, App  
Sequence 122663, A  
Sequence 20619, A  
Sequence 131, App  
Sequence 131, App  
Sequence 131, App  
Sequence 199, App  
Sequence 199, App  
Sequence 204, App  
Sequence 64369, A  
Sequence 14748, A  
Sequence 105413, A  
Sequence 21, Appl  
Sequence 193067, A  
Sequence 114193, A  
Sequence 53269, A  
Sequence 5, Appli  
Sequence 8763, Ap  
Sequence 140685, A  
Sequence 49854, A  
Sequence 152921, A  
Sequence 73345, A  
Sequence 49757, A  
Sequence 1494, Ap  
Sequence 1494, Ap  
Sequence 63, Appl  
Sequence 63, Appl  
Sequence 11721, A

16 11.7 154 US-10-437-963-162284  
17 90.5 745 8 US-08-834-666A-6  
18 90.5 745 12 US-10-282-122A-58683  
19 90 1778 14 US-10-238-075-749  
20 89.5 11.5 270 16 US-10-437-963-122263  
21 89.5 11.5 486 15 US-10-369-493-20619  
22 89 943 9 US-09-996-634-131  
23 89 11.4 943 10 US-09-997-182-131  
24 89 11.4 943 10 US-09-997-181-131  
25 89 11.4 943 14 US-10-193-002-199  
26 89 11.4 943 14 US-10-084-843-204  
27 89 3300 12 US-10-282-122A-64369  
28 88.5 11.4 145 16 US-10-437-963-147748  
29 88.5 11.4 191 16 US-10-437-963-105413  
30 88 11.3 354 10 US-09-820-843A-21  
31 88 616 16 US-10-437-963-193067  
32 87.5 11.2 253 16 US-10-437-963-114193  
33 87.5 11.2 628 12 US-10-282-122A-53269  
34 87.5 11.2 892 10 US-09-952-267-5  
35 87 11.2 482 14 US-10-156-761-8763  
36 86 11.0 448 16 US-10-437-963-140685  
37 86 11.0 2457 12 US-10-282-122A-49854  
38 85.5 11.0 191 16 US-10-437-963-152921  
39 85.5 11.0 842 12 US-10-282-122A-73345  
40 85.5 11.0 1862 12 US-10-282-122A-49757  
41 85 10.9 257 10 US-09-880-748-1494  
42 85 10.9 257 12 US-10-293-418-1494  
43 85 10.9 956 13 US-10-121-032-63  
44 85 10.9 956 14 US-10-093-037-63  
45 84.5 10.8 271 14 US-10-156-761-11721

#### ALIGNMENTS

#### RESULT 1

US-09-741-873B-4  
; Sequence 4, Application US/09741873B  
; Publication NO. US20020081722A1  
; GENERAL INFORMATION:  
; APPLICANT: Normark, Staffan  
; APPLICANT: Olsen, Arne  
; TITLE OF INVENTION: Fibronectin Binding Protein As Well As Its Preparation  
; FILE REFERENCE: 012889-084  
; CURRENT APPLICATION NUMBER: US/09/741, 873B  
; PRIOR FILING DATE: 2003-04-04  
; PRIOR APPLICATION NUMBER: SE 8801723-1  
; PRIOR FILING DATE: 1998-05-06  
; PRIOR APPLICATION NUMBER: US 08/978,878  
; PRIOR FILING DATE: 1997-11-26  
; PRIOR APPLICATION NUMBER: US 07/347,189  
; PRIOR FILING DATE: 1989-05-04  
; PRIOR APPLICATION NUMBER: US 07/789,437  
; PRIOR FILING DATE: 1991-11-06  
; PRIOR APPLICATION NUMBER: US 07/970,846  
; PRIOR FILING DATE: 1992-11-03  
; PRIOR APPLICATION NUMBER: US 08/187,865  
; PRIOR FILING DATE: 1994-01-28  
; PRIOR APPLICATION NUMBER: US 08/318,519  
; PRIOR FILING DATE: 1994-10-05  
; NUMBER OF SEQ ID NOS: 10  
; SOFTWARE: PatentIn version 3.0  
; SEQ ID NO 4  
; LENGTH: 151  
; TYPE: PRT  
; ORGANISM: Escherichia coli  
US-09-741-873B-4

Query Match 66.4%; Score 517; DB 12; Length 151;  
Best Local Similarity 66.2%; Pred. No. 1.7e-43;  
Matches 100; Conservative 22; Mismatches 29; Indels 0; Gaps 0;  
QY 1 MKLLKVAFAAIVVSSGSLAGVVPQWGGGNGGNGSSGPDSTLSIYQVGSANALALQ 60

Db 1 MLLKVAIAAIVFSGSAVAVPQYGGGNGHGGGNSGPNSELNIYQYGGNSALALQ 60  
Qy 61 SDARKSETTITQSGYNGADVGQADNSTIELTQNGFRNNATIDQWNAKNSDITVQYGG 120  
Db 61 TDARNSDLTITQHGNGGADVGQSDSSIDLQTRGFNSATLDQWNGKNSMTVQVFGG 120  
Qy 121 NNAALVNDQLVTRVTHVAHANNATANQY 151  
Db 121 GNGAAVDQTASNSVNVTVQVFGNNATAHQY 151

## RESULT 2

US-09-741-873B-4  
; Sequence 4, Application US/09741873B  
; Publication No. US20040096965A9  
; GENERAL INFORMATION:  
; APPLICANT: Normark, Staffan  
; APPLICANT: Olsen, Arne  
; TITLE OF INVENTION: Fibronectin Binding Protein As Well As Its Preparation  
; FILE REFERENCE: 012889-084  
; CURRENT APPLICATION NUMBER: US/09/741,873B  
; CURRENT FILING DATE: 2003-04-04  
; PRIOR APPLICATION NUMBER: SE 8801723-1  
; PRIOR FILING DATE: 1998-05-06  
; PRIOR APPLICATION NUMBER: US 08/978,878  
; PRIOR FILING DATE: 1997-11-26  
; PRIOR APPLICATION NUMBER: US 07/347,189  
; PRIOR FILING DATE: 1989-05-04  
; PRIOR APPLICATION NUMBER: US 07/789,437  
; PRIOR FILING DATE: 1991-11-06  
; PRIOR APPLICATION NUMBER: US 07/970,846  
; PRIOR FILING DATE: 1992-11-03  
; PRIOR APPLICATION NUMBER: US 08/318,519  
; PRIOR FILING DATE: 1994-01-28  
; NUMBER OF SEQ ID NOS: 10  
; SOFTWARE: PatentIn version 3.0  
; SEQ ID NO 4  
; LENGTH: 151  
; TYPE: PRT  
; ORGANISM: Escherichia coli

## US-09-741-873B-4

Query Match 66.4%; Score 517; DB 12; Length 151;  
Best Local Similarity 66.2%; Pred. No. 1.7e-43;  
Matches 100; Conservative 22; Mismatches 29; Indels 0; Gaps 0;  
Qy 1 MLLKVAIAAIVFSGSAVAVPQYGGGNGHGGGNSGPNSELNIYQYGGNSALALQ 60  
Db 1 MLLKVAIAAIVFSGSAVAVPQYGGGNGHGGGNSGPNSELNIYQYGGNSALALQ 60  
Qy 61 SDARKSETTITQSGYNGADVGQADNSTIELTQNGFRNNATIDQWNAKNSDITVQYGG 120  
Db 61 TDARNSDLTITQHGNGGADVGQSDSSIDLQTRGFNSATLDQWNGKNSMTVQVFGG 120  
Qy 121 NNAALVNDQLVTRVTHVAHANNATANQY 151  
Db 121 GNGAAVDQTASNSVNVTVQVFGNNATAHQY 151

## RESULT 3

US-09-741-873B-2  
; Sequence 2, Application US/09741873B  
; Publication No. US20020081722A1  
; GENERAL INFORMATION:  
; APPLICANT: Normark, Staffan  
; APPLICANT: Olsen, Arne  
; TITLE OF INVENTION: Fibronectin Binding Protein As Well As Its Preparation  
; FILE REFERENCE: 012889-084  
; CURRENT APPLICATION NUMBER: US/09/741,873B  
; CURRENT FILING DATE: 2003-04-04

## US-09-741-873B-2

Query Match 66.4%; Score 439; DB 12; Length 131;  
Best Local Similarity 62.6%; Pred. No. 8e-36;  
Matches 82; Conservative 21; Mismatches 28; Indels 0; Gaps 0;  
Qy 21 GVPQWGGGNGHGGNSGPDSTLSIYQYGSANAALQSDARKSETTITQSGYNGAD 80  
Db 1 GVPQYGGGNGHGGGNSGENSELNTYQYGGNSALALQTDARNSDLTITQHGNGAD 60  
Qy 81 VQGGADNSTIELTQNGFRNNATIDQWNAKNSDITVQYGGNSALALQVTRVTHM 140  
Db 61 VQGSDDSSIDLQTRGFNSATLDQWNGKNSMTVQVFGGNGAAVDQTASNSVNVTVQ 120  
Qy 141 AHANNATANQY 151  
Db 121 GFGNNATAHQY 131

## RESULT 4

US-09-741-873B-2  
; Sequence 2, Application US/09741873B  
; Publication No. US20040096965A9  
; GENERAL INFORMATION:  
; APPLICANT: Normark, Staffan  
; APPLICANT: Olsen, Arne  
; TITLE OF INVENTION: Fibronectin Binding Protein As Well As Its Preparation  
; FILE REFERENCE: 012889-084  
; CURRENT APPLICATION NUMBER: US/09/741,873B  
; CURRENT FILING DATE: 2003-04-04  
; PRIOR APPLICATION NUMBER: SE 8801723-1  
; PRIOR FILING DATE: 1998-05-06  
; PRIOR APPLICATION NUMBER: US 08/978,878  
; PRIOR FILING DATE: 1997-11-26  
; PRIOR APPLICATION NUMBER: US 07/347,189  
; PRIOR FILING DATE: 1989-05-04  
; PRIOR APPLICATION NUMBER: US 07/789,437  
; PRIOR FILING DATE: 1991-11-06  
; PRIOR APPLICATION NUMBER: US 07/970,846  
; PRIOR FILING DATE: 1992-11-03  
; PRIOR APPLICATION NUMBER: US 08/187,865  
; PRIOR FILING DATE: 1994-01-28  
; PRIOR APPLICATION NUMBER: US 08/318,519  
; PRIOR FILING DATE: 1994-10-05  
; NUMBER OF SEQ ID NOS: 10  
; SOFTWARE: PatentIn version 3.0  
; SEQ ID NO 2  
; LENGTH: 131  
; TYPE: PRT  
; ORGANISM: Escherichia coli

## US-09-741-873B-2

Query Match 56.4%; Score 439; DB 12; Length 131;  
Best Local Similarity 56.4%; Score 439; DB 12; Length 131;  
Matches 82; Conservative 21; Mismatches 28; Indels 0; Gaps 0;  
Qy 21 GVPQWGGGNGHGGNSGPDSTLSIYQYGSANAALQSDARKSETTITQSGYNGAD 80  
Db 1 GVPQYGGGNGHGGGNSGENSELNTYQYGGNSALALQTDARNSDLTITQHGNGAD 60  
Qy 81 VQGGADNSTIELTQNGFRNNATIDQWNAKNSDITVQYGGNSALALQVTRVTHM 140  
Db 61 VQGSDDSSIDLQTRGFNSATLDQWNGKNSMTVQVFGGNGAAVDQTASNSVNVTVQ 120  
Qy 141 AHANNATANQY 151  
Db 121 GFGNNATAHQY 131

## US-09-741-873B-2

Query Match 56.4%; Score 439; DB 12; Length 131;

```
Best Local Similarity 62.6%; Pred. No. 8e-36;
Matches 82; Conservative 21; Mismatches 28; Indels 0; Gaps 0;

QY 21 GVPQGGGNNHGGNSGPDSTLSIYQYGSANAALQSDARKSETTITQSGVGNAD 80
      |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 1 GVPVQGGGNNHGGNSGPNSELNTYQYGGGNSALALQTDARNSDLTITQGGGNGAD 60
      |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 81 VGGADNSTIELTQNGFRNATIDQWNAKNSDITVQYGGNNAALVNYDQLVTRVVTHEM 140
      |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 61 VGGSDSDSIDTQRFNGNSAILDQWNGKNSMTVKQFGGNGAAVDQTASNSSVNVTVQ 120
      |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 141 AHANNATANQY 151
      |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 121 GFGNNTAHQY 131
      |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:

RESULT 5
US-10-369-493-20638
; Sequence 20638, Application US/10369493
; Publication No. US20030233675A1
; GENERAL INFORMATION:
; APPLICANT: Cao, Yongwei
; APPLICANT: Hinkle, Gregory J.
; APPLICANT: Slater, Steven C.
; APPLICANT: Goldman, Barry S.
; APPLICANT: Chen, Xianfeng
; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
; TITLE OF INVENTION: PLANTS WITH IMPROVED PROPERTIES
; FILE REFERENCE: 38-10(52052)B
; CURRENT APPLICATION NUMBER: US/10/369,493
; PRIOR FILING DATE: 2003-02-28
; PRIOR APPLICATION NUMBER: US 60/360,039
; PRIOR FILING DATE: 2002-02-21
; NUMBER OF SEQ ID NOS: 47374
; SEQ ID NO 20638
; LENGTH: 445
; TYPE: PRT
; ORGANISM: Rhodospseudomonas palustris
; FEATURE:
; NAME/KEY: unsure
; LOCATION: (1)...(445)
; OTHER INFORMATION: unsure at all Xaa locations
US-10-369-493-20638

Query Match 12.8%; Score 100; DB 15; Length 445;
Best Local Similarity 26.6%; Pred. No. 0.18;
Matches 47; Conservative 20; Mismatches 56; Indels 54; Gaps 8;

QY 7 AAFAA-----IVVGSALAGVVPQGGG-----NHNGG-----GNSSGPDSTLSIYQY 50
      |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 19 AAFADNTVILNQTGNDQQAINTQSGNGSVGAFNGSGLQENGTLSCA-NLLIVKQS 77
      |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 51 GSANAALALQSDARKSETTITQSGYNGADVGGQADNSTIELTQNGFRNATIDQWNAKN 110
      |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 78 GNSNSV-----GRDIQKQSGAGNSAAIFQEGTGSDELQQTGTSNGAVSPSGWNWTN 129
      |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 111 -----SDITVQYGGNNAALVNYDQLVTRVVTHEMAHANNATA-NQ 150
      |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 130 DPGVFNKITQDSSNSGSKSVIQDGKNN-----VFSIKQNTGNSTNSVNO 174
      |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:

RESULT 6
US-10-425-114-49960
; Sequence 49960, Application US/10425114
; Publication No. US20040034888A1
; GENERAL INFORMATION:
; APPLICANT: Liu, Jingdong
; APPLICANT: Zhou, Yihua
; APPLICANT: Kovalic, David K.
; APPLICANT: Screen, Steven E
; APPLICANT: Tabaska, Jack E
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
```

```
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(53313)B
; CURRENT APPLICATION NUMBER: US/10/425,114
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 73128
; SEQ ID NO 49960
; LENGTH: 263
; TYPE: PRT
; ORGANISM: Zea mays
; FEATURE:
; OTHER INFORMATION: Clone ID: 700071884_FLI.pep
US-10-425-114-49960

Query Match 12.5%; Score 97.5; DB 12; Length 263;
Best Local Similarity 26.6%; Pred. No. 0.17;
Matches 34; Conservative 18; Mismatches 53; Indels 23; Gaps 5;

QY 17 SALAGVVPQGGGNNHGGNSGPDSTLSIYQYGSANAALQSDARKSETTITQSG-- 74
      |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 84 SSIAG-----GGGGGGGGGTNGSGSGSGSGSSTAASGSPSGNVADAEGKGAG 138
      |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 75 -----YNGGA--DVGGQADNSTIEL--TQNGFRNATIDQWNAKNSDITVQYGG 120
      |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 139 GGMGGGANGAYGSGAGGKGEGVSGVALAPSSNGYNGGADATGGGSG--AGGGHGG 196
      |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 121 NNRAALVNY 128
      |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 197 GAAGAPSY 204
      |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:

RESULT 7
US-10-282-122A-64573
; Sequence 64573, Application US/10282122A
; Publication No. US20040029129A1
; GENERAL INFORMATION:
; APPLICANT: Wang, Liangsu
; APPLICANT: Zamudio, Carlos
; APPLICANT: Malone, Cheryl
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari
; APPLICANT: Zyskind, Judith
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John
; APPLICANT: Carr, Grant
; APPLICANT: Yamamoto, Robert
; APPLICANT: Forsyth, R.
; APPLICANT: Xu, H.
; TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
; FILE REFERENCE: ELITRA.034A
; CURRENT APPLICATION NUMBER: US/10/282,122A
; CURRENT FILING DATE: 2003-02-20
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/230,335
; PRIOR FILING DATE: 2000-09-06
; PRIOR APPLICATION NUMBER: 60/230,347
; PRIOR FILING DATE: 2000-09-09
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/267,636
; PRIOR FILING DATE: 2001-02-09
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 78614
```



RESULT 13  
US-09-972-115A-8  
Sequence 8, Application US/099721115A  
Publication No. US20030032769A1  
GENERAL INFORMATION:  
APPLICANT: Geron Corporation  
APPLICANT: Gregg, Morin B.  
APPLICANT: Walter, Funk D.  
APPLICANT: Misczynslaw, Piatyszek A  
TITLE OF INVENTION: A Second Mammal  
FILE REFERENCE: 080/003C  
CURRENT APPLICATION NUMBER: US/09/972-115A-8  
CURRENT FILING DATE: 2001-10-05  
PRIOR APPLICATION NUMBER: US 66/128,000



GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: August 2, 2004, 14:48:33 ; Search time 167.9 Seconds  
(without alignments)  
877.809 Million cell updates/sec

Title: US-09-543-407-12

Perfect score: 779

Sequence: 1 MKLLKVAAPAAIVVSGSALA.....VTRVTVTHEVAHANNATANCY 151

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 6019581 seqs, 976053577 residues

Total number of hits satisfying chosen parameters: 6019581

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

- Pending Patents\_AA\_Main:
- 1: /cgn2\_6/ptodata/2/paa/PCTUS\_COMB.pep.\*
  - 2: /cgn2\_6/ptodata/2/paa/US06\_COMB.pep.\*
  - 3: /cgn2\_6/ptodata/2/paa/US07\_COMB.pep.\*
  - 4: /cgn2\_6/ptodata/2/paa/US08\_COMB.pep.\*
  - 5: /cgn2\_6/ptodata/2/paa/US081\_COMB.pep.\*
  - 6: /cgn2\_6/ptodata/2/paa/US082\_COMB.pep.\*
  - 7: /cgn2\_6/ptodata/2/paa/US083\_COMB.pep.\*
  - 8: /cgn2\_6/ptodata/2/paa/US084\_COMB.pep.\*
  - 9: /cgn2\_6/ptodata/2/paa/US085\_COMB.pep.\*
  - 10: /cgn2\_6/ptodata/2/paa/US086\_COMB.pep.\*
  - 11: /cgn2\_6/ptodata/2/paa/US087\_COMB.pep.\*
  - 12: /cgn2\_6/ptodata/2/paa/US088\_COMB.pep.\*
  - 13: /cgn2\_6/ptodata/2/paa/US089\_COMB.pep.\*
  - 14: /cgn2\_6/ptodata/2/paa/US090\_COMB.pep.\*
  - 15: /cgn2\_6/ptodata/2/paa/US091\_COMB.pep.\*
  - 16: /cgn2\_6/ptodata/2/paa/US092\_COMB.pep.\*
  - 17: /cgn2\_6/ptodata/2/paa/US093\_COMB.pep.\*
  - 18: /cgn2\_6/ptodata/2/paa/US094\_COMB.pep.\*
  - 19: /cgn2\_6/ptodata/2/paa/US095\_COMB.pep.\*
  - 20: /cgn2\_6/ptodata/2/paa/US096\_COMB.pep.\*
  - 21: /cgn2\_6/ptodata/2/paa/US097A\_COMB.pep.\*
  - 22: /cgn2\_6/ptodata/2/paa/US097B\_COMB.pep.\*
  - 23: /cgn2\_6/ptodata/2/paa/US098\_COMB.pep.\*
  - 24: /cgn2\_6/ptodata/2/paa/US099A\_COMB.pep.\*
  - 25: /cgn2\_6/ptodata/2/paa/US099B\_COMB.pep.\*
  - 26: /cgn2\_6/ptodata/2/paa/US100\_COMB.pep.\*
  - 27: /cgn2\_6/ptodata/2/paa/US101\_COMB.pep.\*
  - 28: /cgn2\_6/ptodata/2/paa/US102\_COMB.pep.\*
  - 29: /cgn2\_6/ptodata/2/paa/US103\_COMB.pep.\*
  - 30: /cgn2\_6/ptodata/2/paa/US104\_COMB.pep.\*
  - 31: /cgn2\_6/ptodata/2/paa/US106\_COMB.pep.\*
  - 32: /cgn2\_6/ptodata/2/paa/US107\_COMB.pep.\*
  - 33: /cgn2\_6/ptodata/2/paa/US60\_COMB.pep.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	ID	Description
------------	-------	-------	--------	----	-------------

1	779	100.0	151	19	US-09-543-407-12	Sequence 12, Appl
2	696	89.3	151	19	US-09-543-407-14	Sequence 14, Appl
3	695	89.2	151	19	US-09-543-407-5	Sequence 5, Appl
4	690	88.6	151	6	US-08-233-642A-57	Sequence 57, Appl
5	655	84.1	151	19	US-09-543-407-24	Sequence 24, Appl
6	614	78.8	151	19	US-09-543-407-26	Sequence 26, Appl
7	613	78.7	151	19	US-09-543-407-18	Sequence 18, Appl
8	611	78.4	151	19	US-09-543-407-20	Sequence 20, Appl
9	608	78.0	131	19	US-09-543-407-31	Sequence 31, Appl
10	605	77.7	151	19	US-09-543-407-28	Sequence 28, Appl
11	604	77.5	151	19	US-09-543-407-22	Sequence 22, Appl
12	603	77.4	151	19	US-09-543-407-30	Sequence 30, Appl
13	580	74.5	151	19	US-09-543-407-16	Sequence 16, Appl
14	560	71.9	120	6	US-08-233-642A-55	Sequence 55, Appl
15	520	66.8	151	19	US-09-543-407-4	Sequence 4, Appl
16	517	66.4	151	13	US-08-978-878-7	Sequence 7, Appl
17	515	66.1	151	33	US-60-352-946-2	Sequence 4, Appl
18	515	66.1	151	33	US-60-444-371-2	Sequence 2, Appl
19	481	61.7	158	16	US-09-252-691-5834	Sequence 5834, Ap
20	481	61.7	158	16	US-09-252-691C-5834	Sequence 5834, Ap
21	481	61.7	158	30	US-10-417-886-5834	Sequence 5834, Ap
22	476	61.1	109	19	US-09-543-407-34	Sequence 34, Appl
23	439	56.4	131	13	US-08-978-878-2	Sequence 2, Appl
24	439	56.4	131	21	US-09-741-873B-2	Sequence 37, Appl
25	347	44.5	68	19	US-09-543-407-35	Sequence 35, Appl
26	335	43.0	109	19	US-09-543-407-32	Sequence 32, Appl
27	247.5	31.8	70	19	US-09-543-407-39	Sequence 39, Appl
28	237	30.4	48	19	US-09-543-407-6	Sequence 6, Appl
29	100.5	12.9	151	19	US-09-543-407-6	Sequence 5833, Ap
30	100.5	12.9	186	16	US-09-252-691-5833	Sequence 5833, Ap
31	100.5	12.9	186	16	US-09-252-691C-5833	Sequence 5833, Ap
32	100.5	12.9	186	30	US-10-417-886-5833	Sequence 5833, Ap
33	100.5	12.9	186	30	US-10-417-886-5833	Sequence 20638, A
34	100	12.8	445	29	US-10-369-493-20638	Sequence 20638, A
35	100	12.8	445	33	US-60-360-039-20638	Sequence 358, App
36	100	12.8	1249	30	US-10-455-719-358	Sequence 357, App
37	100	12.8	1249	33	US-60-385-568-357	Sequence 358, App
38	100	12.8	1249	33	US-60-446-775-358	Sequence 9317, Ap
39	100	12.8	1266	27	US-10-179-131-9317	Sequence 24084, A
40	98.5	12.6	850	20	US-09-614-150-24084	Sequence 24084, A
41	98.5	12.6	850	20	US-09-614-150A-24084	Sequence 24179, A
42	98.5	12.6	850	33	US-60-191-637-24179	Sequence 19019, A
43	98.5	12.6	850	33	US-60-191-681-19019	Sequence 14916, A
44	98.5	12.6	1028	20	US-09-614-150-14916	Sequence 14916, A
45	98.5	12.6	1028	20	US-09-614-150A-14916	Sequence 14916, A

ALIGNMENTS

RESULT 1  
US-09-543-407-12  
; Sequence 12, Application US/09543407  
; GENERAL INFORMATION:  
; APPLICANT: White, Aaron P.  
; APPLICANT: Doran, James L.  
; APPLICANT: Collinson, S. Karen  
; APPLICANT: Kay, William W.  
; TITLE OF INVENTION: BACTERIAL FIBRIL SYSTEM FOR  
; TITLE OF INVENTION: PRESENTATION OF HETEROLOGOUS PEPTIDE SEQUENCES  
; FILE REFERENCE: 920043.406  
; CURRENT APPLICATION NUMBER: US/09/543,407  
; CURRENT FILING DATE: 2000-04-05  
; NUMBER OF SEQ ID NOS: 59  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 12  
; LENGTH: 151  
; TYPE: PRT  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: Recombinant Salmonella enteritidis 3b afgA  
; OTHER INFORMATION: sequence containing the replacement fragment  
; OTHER INFORMATION: encoding PT3 from GP63 of Leishmania major.

US-09-543-407-12

Query Match 100.0%; Score 779; DB 19; Length 151;  
Best Local Similarity 100.0%; Pred. No. 2.9e-74;  
Matches 151; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MKLLKVAFAAIVVSGSALAGVVPQWGGGNGHNGGNSGPDSTLSIYQYGSANAALALQ 60  
DB 1 MKLLKVAFAAIVVSGSALAGVVPQWGGGNGHNGGNSGPDSTLSIYQYGSANAALALQ 60

QY 61 SDARKSETTITQSGYNGADVGQADNSTIELTQNGFRNNATIDQWNAKNSDITVGYGG 120  
DB 61 SDARKSETTITQSGYNGADVGQADNSTIELTQNGFRNNATIDQWNAKNSDITVGYGG 120

QY 121 NNAALVNYDQLVTRVWTHEMAHANNATANY 151  
DB 121 NNAALVNYDQLVTRVWTHEMAHANNATANY 151

RESULT 2

US-09-543-407-14  
; Sequence 14, Application US/09543407  
; GENERAL INFORMATION:  
; APPLICANT: White, Aaron P.  
; APPLICANT: Doran, James L.  
; APPLICANT: Collinson, S. Karen  
; APPLICANT: Kay, William W.  
; TITLE OF INVENTION: BACTERIAL FIMBRIAL SYSTEM FOR  
; PRESENTATION OF HETEROLOGOUS PEPTIDE SEQUENCES  
; FILE REFERENCE: 920043.406  
; CURRENT APPLICATION NUMBER: US/09/543.407  
; CURRENT FILING DATE: 2000-04-05  
; NUMBER OF SEQ ID NOS: 59  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 14  
; LENGTH: 151  
; TYPE: PRT  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: Recombinant Salmonella enteritidis 3b afgA  
; sequence containing the replacement fragment  
; OTHER INFORMATION: encoding P73 from GP63 of Leishmania major.

US-09-543-407-14

Query Match 89.3%; Score 696; DB 19; Length 151;  
Best Local Similarity 87.6%; Pred. No. 2e-65;  
Matches 141; Conservative 0; Mismatches 20; Indels 20; Gaps 2;

QY 1 MKLLKVAFAAIVVSGSALAGVVPQWGGGNGHNGGNSGPDSTLSIYQYGSANAALALQ 60  
DB 1 MKLLKVAFAAIVVSGSALAGVVPQWGGGNGHNGGNSGPDSTLSIYQYGSANAALALQ 60

QY 61 SDARKSETTITQSGYNGADVGQADNSTIELTQNGFRNNATIDQWNAKNSDITVGYGG 120  
DB 61 SDARKSETTITQSGYNGADVGQADNSTIELTQNGFRNNATIDQWNAKNSDITVGYGG 117

QY 121 NNAALVNYDQLVTRVWTHEMAHANNATANY 151  
DB 118 -----YDQLVTRVWTHEMAHANSVMVRQVGFNNATANY 151

RESULT 3

US-09-543-407-5  
; Sequence 5, Application US/09543407  
; GENERAL INFORMATION:  
; APPLICANT: White, Aaron P.  
; APPLICANT: Doran, James L.  
; APPLICANT: Collinson, S. Karen  
; APPLICANT: Kay, William W.  
; TITLE OF INVENTION: BACTERIAL FIMBRIAL SYSTEM FOR  
; PRESENTATION OF HETEROLOGOUS PEPTIDE SEQUENCES  
; FILE REFERENCE: 920043.406  
; CURRENT APPLICATION NUMBER: US/09/543.407

; CURRENT FILING DATE: 2000-04-05  
; NUMBER OF SEQ ID NOS: 59  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 5  
; LENGTH: 151  
; TYPE: PRT  
; ORGANISM: Salmonella enteritidis  
US-09-543-407-5

Query Match 89.2%; Score 695; DB 19; Length 151;  
Best Local Similarity 90.1%; Pred. No. 2.5e-65;  
Matches 136; Conservative 4; Mismatches 11; Indels 0; Gaps 0;

QY 1 MKLLKVAFAAIVVSGSALAGVVPQWGGGNGHNGGNSGPDSTLSIYQYGSANAALALQ 60  
DB 1 MKLLKVAFAAIVVSGSALAGVVPQWGGGNGHNGGNSGPDSTLSIYQYGSANAALALQ 60

QY 61 SDARKSETTITQSGYNGADVGQADNSTIELTQNGFRNNATIDQWNAKNSDITVGYGG 120  
DB 61 SDARKSETTITQSGYNGADVGQADNSTIELTQNGFRNNATIDQWNAKNSDITVGYGG 120

QY 121 NNAALVNYDQLVTRVWTHEMAHANNATANY 151  
DB 121 NNAALVNYDQLVTRVWTHEMAHANNATANY 151

RESULT 4

US-08-233-642A-57  
; Sequence 57, Application US/08233642A  
; GENERAL INFORMATION:  
; APPLICANT: Kay, William W.  
; APPLICANT: Collinson, S. Karen  
; APPLICANT: Clouthier, Sharon C.  
; APPLICANT: Doran, James L.  
; TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR SALMONELLA-  
; BASED VACCINES  
; NUMBER OF SEQUENCES: 58  
; NUMBER OF SEQUENCES: 58  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Seed and Berry  
; STREET: 6300 Columbia Center, 701 Fifth Avenue  
; CITY: Seattle  
; STATE: Washington  
; COUNTRY: U.S.A.  
; ZIP: 98104-7092  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/233.642A  
; FILING DATE: 26-APR-1994  
; CLASSIFICATION: 424  
; ATTORNEY/AGENT INFORMATION:  
; NAME: King, Joshua  
; REGISTRATION NUMBER: 35,570  
; REFERENCE/DOCKET NUMBER: 920043.403C3  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (206) 622-4900  
; TELEFAX: (206) 682-6031  
; TELEX: 3723836 SEEDANBERRY  
; INFORMATION FOR SEQ ID NO: 57:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 151 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
US-08-233-642A-57

Query Match 88.6%; Score 690; DB 6; Length 151;  
Best Local Similarity 89.4%; Pred. No. 8.6e-65;  
Matches 135; Conservative 4; Mismatches 12; Indels 0; Gaps 0;



```
QY 1 MKLLKVAFAAIVVSGSALAGVVPQWGGGNGHNGSGPDSLTSLIYQYGSANAALALQ 60
Db 1 MKLLKVAFAAIVVSGSALAGVVPQWGGGNGHNGSGPDSLTSLIYQYGSANAALALQ 60
QY 61 SDARKSETTITQSGYNGADVCGGADNSTIETQNGFRNNATIDOWNAKNSDITVGYGG 120
Db 61 SDARKSETTITQSGYNGADVCGGADNSTIETQNGFRNNATIDOWNAKNSDITVGYGG 120
QY 121 NNAALVNYDQLVTRVVTHEMAHANNATANQY 151
Db 121 NNPALVNQTASDSSVMVRQVGFGNATANQY 151

RESULT 5
US-09-543-407-24
; Sequence 24, Application US/09543407
; GENERAL INFORMATION:
; APPLICANT: White, Aaron P.
; APPLICANT: Doran, James L.
; APPLICANT: Collinson, S. Karen
; TITLE OF INVENTION: BACTERIAL FIMBRIAL SYSTEM FOR
; TITLE OF INVENTION: PRESENTATION OF HETEROLOGOUS PEPTIDE SEQUENCES
; FILE REFERENCE: 920043.406
; CURRENT APPLICATION NUMBER: US/09/543,407
; CURRENT FILING DATE: 2000-04-05
; NUMBER OF SEQ ID NOS: 59
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 24
; LENGTH: 151
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Recombinant Salmonella enteritidis 3b afgA
; OTHER INFORMATION: sequence containing the replacement fragment
; OTHER INFORMATION: encoding PT3 from GP63 of Leishmania major.
US-09-543-407-24

Query Match 84.1%; Score 655; DB 19; Length 151;
Best Local Similarity 79.8%; Pred. No. 4.6e-61;
Matches 134; Conservative 0; Mismatches 0; Indels 34; Gaps 2;

QY 1 MKLLKVAFAAIVVSGSALAGVVPQWGGGNGHNGSGPDSLTSLIYQYGSANAALALQ 60
Db 1 MKLLKVAFAAIVVSGSALAGVVPQWGGGNGHNGSGPDSLTSLIYQYGSANAALALQ 60
QY 61 SDARKSETTITQSGYNGADVCGGADNSTIETQNGFRNNATIDOWNAKNSDITVGYGG 120
Db 61 SDARKSETTITQSGYNGADVCGGADNSTIETQNGFRNNATIDOWNAKNSDITVGYGG 120
QY 121 NNAALVNYDQLVTRVVTHEMAHANNATANQY 151
Db 110 -----NYDQLVTRVVTHEMAHANNATANQY 151

RESULT 6
US-09-543-407-26
; Sequence 26, Application US/09543407
; GENERAL INFORMATION:
; APPLICANT: White, Aaron P.
; APPLICANT: Doran, James L.
; APPLICANT: Collinson, S. Karen
; TITLE OF INVENTION: BACTERIAL FIMBRIAL SYSTEM FOR
; TITLE OF INVENTION: PRESENTATION OF HETEROLOGOUS PEPTIDE SEQUENCES
; FILE REFERENCE: 920043.406
; CURRENT APPLICATION NUMBER: US/09/543,407
; CURRENT FILING DATE: 2000-04-05
; NUMBER OF SEQ ID NOS: 59
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 26
; LENGTH: 151
```

```
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Recombinant Salmonella enteritidis 3b afgA
; OTHER INFORMATION: sequence containing the replacement fragment
; OTHER INFORMATION: encoding PT3 from GP63 of Leishmania major.
US-09-543-407-26

Query Match 78.8%; Score 614; DB 19; Length 151;
Best Local Similarity 80.8%; Pred. No. 1.1e-56;
Matches 122; Conservative 6; Mismatches 23; Indels 0; Gaps 0;

QY 1 MKLLKVAFAAIVVSGSALAGVVPQWGGGNGHNGSGPDSLTSLIYQYGSANAALALQ 60
Db 1 MKLLKVAFAAIVVSGSALAGVVPQWGGGNGHNGSGPDSLTSLIYQYGSANAALYDQ 60
QY 61 SDARKSETTITQSGYNGADVCGGADNSTIETQNGFRNNATIDOWNAKNSDITVGYGG 120
Db 61 LVTRVVTHEMAHAGYNGADVCGGADNSTIETQNGFRNNATIDOWNAKNSDITVGYGG 120
QY 121 NNAALVNYDQLVTRVVTHEMAHANNATANQY 151
Db 121 NNAALVNQTASDSSVMVRQVGFGNATANQY 151

RESULT 7
US-09-543-407-18
; Sequence 18, Application US/09543407
; GENERAL INFORMATION:
; APPLICANT: White, Aaron P.
; APPLICANT: Doran, James L.
; APPLICANT: Collinson, S. Karen
; TITLE OF INVENTION: BACTERIAL FIMBRIAL SYSTEM FOR
; TITLE OF INVENTION: PRESENTATION OF HETEROLOGOUS PEPTIDE SEQUENCES
; FILE REFERENCE: 920043.406
; CURRENT APPLICATION NUMBER: US/09/543,407
; CURRENT FILING DATE: 2000-04-05
; NUMBER OF SEQ ID NOS: 59
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 18
; LENGTH: 151
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Recombinant Salmonella enteritidis 3b afgA
; OTHER INFORMATION: sequence containing the replacement fragment
; OTHER INFORMATION: encoding PT3 from GP63 of Leishmania major.
US-09-543-407-18

Query Match 78.7%; Score 613; DB 19; Length 151;
Best Local Similarity 80.8%; Pred. No. 1.4e-56;
Matches 122; Conservative 7; Mismatches 22; Indels 0; Gaps 0;

QY 1 MKLLKVAFAAIVVSGSALAGVVPQWGGGNGHNGSGPDSLTSLIYQYGSANAALALQ 60
Db 1 MKLLKVAFAAIVVSGSALAGVVPQWGGGNGHNGSGPDSLTSLIYQYGSANAALQ 60
QY 61 SDARKSETTITQSGYNGADVCGGADNSTIETQNGFRNNATIDOWNAKNSDITVGYGG 120
Db 61 SDARKSETTITQSGYNGADVCGGADNSTIETQNGFRNNATIDOWNAKNSDITVGYGG 120
QY 121 NNAALVNYDQLVTRVVTHEMAHANNATANQY 151
Db 121 NNAALVNQTASDSSVMVRQVGFGNATANQY 151

RESULT 8
US-09-543-407-20
; Sequence 20, Application US/09543407
; GENERAL INFORMATION:
; APPLICANT: White, Aaron P.
; APPLICANT: Doran, James L.
```

APPLICANT: Collinson, S. Karen  
APPLICANT: Kay, William W.  
TITLE OF INVENTION: BACTERIAL FIMBRIAL SYSTEM FOR  
FILE REFERENCE: 920043.406  
CURRENT APPLICATION NUMBER: US/09/543.407  
CURRENT FILING DATE: 2000-04-05  
NUMBER OF SEQ ID NOS: 59  
SOFTWARE: FastSeq for Windows Version 4.0  
SEQ ID NO 20  
LENGTH: 151  
TYPE: PRT  
ORGANISM: Artificial Sequence  
FEATURE:  
OTHER INFORMATION: Recombinant Salmonella enteritidis 3b afga  
OTHER INFORMATION: sequence containing the replacement fragment  
OTHER INFORMATION: encoding PT3 from GP63 of Leishmania major.  
US-09-543-407-20

Query Match 78.4%; Score 611; DB 19; Length 151;  
Best Local Similarity 80.8%; Pred. No. 2.2e-56;  
Matches 122; Conservative 7; Mismatches 22; Indels 0; Gaps 0;  
QY 1 MLLKVAFAAIVVSGSALAGVVPQWGGGNGHNGSGDPDSTLSIYQYGSANAALQ 60  
DB 1 MLLKVAFAAIVVSGSALAGVVPQWGGGNGHNGSGDPDSTLSIYQYGSANAALQ 60  
QY 61 SDARKSETTITQSGYNGADVGQADNSTIELTQNGFRNNATIDQWNAKNSDITVGYGG 120  
DB 61 SDARKYDQLVTRVVTHEMAHAGQADNSTIELTQNGFRNNATIDQWNAKNSDITVGYGG 120  
QY 121 NNAALVNDQLVTRVVTHEMAHANNATANY 151  
DB 121 NNAALVNDQASDSSVMVRQVGFNNATANY 151

RESULT 9  
US-09-543-407-31

Sequence 31, Application US/09543407  
GENERAL INFORMATION:  
APPLICANT: White, Aaron P.  
APPLICANT: Doran, James L.  
APPLICANT: Collinson, S. Karen  
APPLICANT: Kay, William W.  
TITLE OF INVENTION: BACTERIAL FIMBRIAL SYSTEM FOR  
FILE REFERENCE: 920043.406  
CURRENT APPLICATION NUMBER: US/09/543.407  
CURRENT FILING DATE: 2000-04-05  
NUMBER OF SEQ ID NOS: 59  
SOFTWARE: FastSeq for Windows Version 4.0  
SEQ ID NO 31  
LENGTH: 131  
TYPE: PRT  
ORGANISM: Salmonella enteritidis  
US-09-543-407-31

Query Match 78.0%; Score 608; DB 19; Length 131;  
Best Local Similarity 88.5%; Pred. No. 3.9e-56;  
Matches 116; Conservative 4; Mismatches 11; Indels 0; Gaps 0;  
QY 21 GVVPQWGGGNGHNGSGDPDSTLSIYQYGSANAALQSDARKSETTITQSGYNGAD 80  
DB 1 GVVPQWGGGNGHNGSGDPDSTLSIYQYGSANAALQSDARKSETTITQSGYNGAD 60  
QY 81 VQGGADNSTIELTQNGFRNNATIDQWNAKNSDITVQYGGNNAALVNDQLVTRVVTHEM 140  
DB 61 VQGGADNSTIELTQNGFRNNATIDQWNAKNSDITVQYGGNNAALVNDQASDSSVMVRQV 120  
QY 141 AHANNATANY 151  
DB 121 GFNNATANY 131

## RESULT 10

US-09-543-407-28  
Sequence 28, Application US/09543407  
GENERAL INFORMATION:  
APPLICANT: White, Aaron P.  
APPLICANT: Doran, James L.  
APPLICANT: Collinson, S. Karen  
APPLICANT: Kay, William W.  
TITLE OF INVENTION: BACTERIAL FIMBRIAL SYSTEM FOR  
FILE REFERENCE: 920043.406  
CURRENT APPLICATION NUMBER: US/09/543.407  
CURRENT FILING DATE: 2000-04-05  
NUMBER OF SEQ ID NOS: 59  
SOFTWARE: FastSeq for Windows Version 4.0  
SEQ ID NO 28  
LENGTH: 151  
TYPE: PRT  
ORGANISM: Artificial Sequence  
FEATURE:  
OTHER INFORMATION: Recombinant Salmonella enteritidis 3b afga  
OTHER INFORMATION: sequence containing the replacement fragment  
OTHER INFORMATION: encoding PT3 from GP63 of Leishmania major.  
US-09-543-407-28

Query Match 77.7%; Score 605; DB 19; Length 151;  
Best Local Similarity 80.1%; Pred. No. 9.8e-56;  
Matches 121; Conservative 5; Mismatches 25; Indels 0; Gaps 0;  
QY 1 MLLKVAFAAIVVSGSALAGVVPQWGGGNGHNGSGDPDSTLSIYQYGSANAALQ 60  
DB 1 MLLKVAFAAIVVSGSALAGVVPQWGGGNGHNGSGDPDSTLSIYQYGSANAALQ 60  
QY 61 SDARKSETTITQSGYNGADVGQADNSTIELTQNGFRNNATIDQWNAKNSDITVGYGG 120  
DB 61 SDARKSETTITQSGYNGADYDQLVTRVVTHEMAHAFNNATIDQWNAKNSDITVGYGG 120  
QY 121 NNAALVNDQLVTRVVTHEMAHANNATANY 151  
DB 121 NNAALVNDQASDSSVMVRQVGFNNATANY 151

## RESULT 11

US-09-543-407-22  
Sequence 22, Application US/09543407  
GENERAL INFORMATION:  
APPLICANT: White, Aaron P.  
APPLICANT: Doran, James L.  
APPLICANT: Collinson, S. Karen  
APPLICANT: Kay, William W.  
TITLE OF INVENTION: BACTERIAL FIMBRIAL SYSTEM FOR  
FILE REFERENCE: 920043.406  
CURRENT APPLICATION NUMBER: US/09/543.407  
CURRENT FILING DATE: 2000-04-05  
NUMBER OF SEQ ID NOS: 59  
SOFTWARE: FastSeq for Windows Version 4.0  
SEQ ID NO 22  
LENGTH: 151  
TYPE: PRT  
ORGANISM: Artificial Sequence  
FEATURE:  
OTHER INFORMATION: Recombinant Salmonella enteritidis 3b afga  
OTHER INFORMATION: sequence containing the replacement fragment  
OTHER INFORMATION: encoding PT3 from GP63 of Leishmania major.  
US-09-543-407-22

Query Match 77.5%; Score 604; DB 19; Length 151;  
Best Local Similarity 80.8%; Pred. No. 1.2e-55;  
Matches 122; Conservative 6; Mismatches 23; Indels 0; Gaps 0;  
QY 1 MLLKVAFAAIVVSGSALAGVVPQWGGGNGHNGSGDPDSTLSIYQYGSANAALQ 60

```
Db 1 MKLLKVAFAAIVVSGSALAGVVPQWGGGNGHNGGSGPDSTLSIYQGSANAALALQ 60
QY 61 SDARKSETTITQSGYNGADVGQGDNSTIELTQNGFRNNATIDOWNAKNSDITVGOYGG 120
Db 61 SDARKSETTITQSGYNGADVGQGDNSTIELTQNGFRNNATIDOWNAKNSDITVGOYGG 120
QY 121 NNAALVNYDQLVTRVVTHEMAHANNATANQY 151
Db 121 NNAALVNYDQLVTRVVTHEMAHANNATANQY 151

RESULT 12
US-09-543-407-30
; Sequence 30, Application US/09543407
; GENERAL INFORMATION:
; APPLICANT: White, Aaron P.
; APPLICANT: Doran, James L.
; APPLICANT: Collinson, S. Karen
; TITLE OF INVENTION: BACTERIAL FIMBRIAL SYSTEM FOR
; PRESENTATION OF HETEROLOGOUS PEPTIDE SEQUENCES
; FILE REFERENCE: 920043.406
; CURRENT FILING DATE: 2000-04-05
; NUMBER OF SEQ ID NOS: 59
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 30
; LENGTH: 151
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Recombinant Salmonella enteritidis 3b afgA
; OTHER INFORMATION: sequence containing the replacement fragment
; OTHER INFORMATION: encoding PT3 from GP63 of Leishmania major.
US-09-543-407-30

Query Match 77.4%; Score 603; DB 19; Length 151;
Best Local Similarity 80.8%; Pred. No. 1.6e-55;
Matches 122; Conservative 6; Mismatches 23; Indels 0; Gaps 0;

QY 1 MKLLKVAFAAIVVSGSALAGVVPQWGGGNGHNGGSGPDSTLSIYQGSANAALALQ 60
Db 1 MKLLKVAFAAIVVSGSALAGVVPQWGGGNGHNGGSGPDSTLSIYQGSANAALALQ 60
QY 61 SDARKSETTITQSGYNGADVGQGDNSTIELTQNGFRNNATIDOWNAKNSDITVGOYGG 120
Db 61 SDARKSETTITQSGYNGADVGQGDNSTIELTQNGFRNNATIDOWNAKNSDITVGOYGG 120
QY 121 NNAALVNYDQLVTRVVTHEMAHANNATANQY 151
Db 121 NNAALVNYDQLVTRVVTHEMAHANNATANQY 151

RESULT 13
US-09-543-407-16
; Sequence 16, Application US/09543407
; GENERAL INFORMATION:
; APPLICANT: White, Aaron P.
; APPLICANT: Doran, James L.
; APPLICANT: Collinson, S. Karen
; APPLICANT: Kay, William W.
; TITLE OF INVENTION: BACTERIAL FIMBRIAL SYSTEM FOR
; PRESENTATION OF HETEROLOGOUS PEPTIDE SEQUENCES
; FILE REFERENCE: 920043.406
; CURRENT FILING DATE: 2000-04-05
; NUMBER OF SEQ ID NOS: 59
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 16
; LENGTH: 151
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Recombinant Salmonella enteritidis 3b afgA
; OTHER INFORMATION: sequence containing the replacement fragment
; OTHER INFORMATION: encoding PT3 from GP63 of Leishmania major.
US-09-543-407-16

Query Match 77.4%; Score 603; DB 19; Length 151;
Best Local Similarity 80.8%; Pred. No. 1.6e-55;
Matches 122; Conservative 6; Mismatches 23; Indels 0; Gaps 0;

QY 1 MKLLKVAFAAIVVSGSALAGVVPQWGGGNGHNGGSGPDSTLSIYQGSANAALALQ 60
Db 1 MKLLKVAFAAIVVSGSALAGVVPQWGGGNGHNGGSGPDSTLSIYQGSANAALALQ 60
QY 61 SDARKSETTITQSGYNGADVGQGDNSTIELTQNGFRNNATIDOWNAKNSDITVGOYGG 120
Db 61 SDARKSETTITQSGYNGADVGQGDNSTIELTQNGFRNNATIDOWNAKNSDITVGOYGG 120
QY 121 NNAALVNYDQLVTRVVTHEMAHANNATANQY 151
Db 121 NNAALVNYDQLVTRVVTHEMAHANNATANQY 151
```

```
; FEATURE:
; OTHER INFORMATION: Recombinant Salmonella enteritidis 3b afgA
; OTHER INFORMATION: sequence containing the replacement fragment
; OTHER INFORMATION: encoding PT3 from GP63 of Leishmania major.
US-09-543-407-16

Query Match 74.5%; Score 580; DB 19; Length 151;
Best Local Similarity 80.1%; Pred. No. 4.5e-53;
Matches 121; Conservative 7; Mismatches 23; Indels 0; Gaps 0;

QY 1 MKLLKVAFAAIVVSGSALAGVVPQWGGGNGHNGGSGPDSTLSIYQGSANAALALQ 60
Db 1 MKLLKVAFAAIVVSGSALAGVVPQWGGGNGHNGGSGPDSTLSIYQGSANAALALQ 60
QY 61 SDARKSETTITQSGYNGADVGQGDNSTIELTQNGFRNNATIDOWNAKNSDITVGOYGG 120
Db 61 SDARKSETTITQSGYNGADVGQGDNSTIELTQNGFRNNATIDOWNAKNSDITVGOYGG 120
QY 121 NNAALVNYDQLVTRVVTHEMAHANNATANQY 151
Db 121 NNAALVNYDQLVTRVVTHEMAHANNATANQY 151

RESULT 14
US-08-233-642A-55
; Sequence 55, Application US/08233642A
; GENERAL INFORMATION:
; APPLICANT: Kay, William W.
; APPLICANT: Collinson, S. Karen
; APPLICANT: Clouthier, Sharon C.
; APPLICANT: Doran, James L.
; TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR SALMONELLA-
; NUMBER OF SEQUENCES: 58
; NUMBER OF SEQUENCES: 58
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Seed and Berry
; STREET: 6300 Columbia Center, 701 Fifth Avenue
; CITY: Seattle
; STATE: Washington
; COUNTRY: U.S.A.
; ZIP: 98104-7092
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/233,642A
; FILING DATE: 26-APR-1994
; CLASSIFICATION: 424
; ATTORNEY/AGENT INFORMATION:
; NAME: King, Joshua
; REGISTRATION NUMBER: 35,570
; REFERENCE/DOCKET NUMBER: 920043.403C3
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (206) 622-4900
; TELEFAX: (206) 682-6031
; TELEX: 3723836 SEEDANBERRY
; INFORMATION FOR SEQ ID NO: 55:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 120 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-233-642A-55

Query Match 71.9%; Score 560; DB 6; Length 120;
Best Local Similarity 100.0%; Pred. No. 4.5e-51;
Matches 106; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 22 VVPQWGGGNGHNGGSGPDSTLSIYQGSANAALALQSDARKSETTITQSGYNGADV 81
```

Db 1 VVPQWGGGNHNGGSSGPDSTLSIYQYGSANAALALQSDARKSETTITQSGYNGADV 60  
Qy 82 GQGADNSTIELTQNGFRNNATIDQWNAKNSDITVGOYGGNNAALVN 127  
Db 61 GQGADNSTIELTQNGFRNNATIDQWNAKNSDITVGOYGGNNAALVN 106

RESULT 15  
US-09-543-407-7  
; Sequence 7, Application US/09543407  
; GENERAL INFORMATION:  
; APPLICANT: White, Aaron P.  
; APPLICANT: Doran, James L.  
; APPLICANT: Collinson, S. Karen  
; APPLICANT: Kay, William W.  
; TITLE OF INVENTION: BACTERIAL FIBRIAL SYSTEM FOR  
; TITLE OF INVENTION: PRESENTATION OF HETEROLOGOUS PEPTIDE SEQUENCES  
; FILE REFERENCE: 920043-406  
; CURRENT APPLICATION NUMBER: US/09/543,407  
; CURRENT FILING DATE: 2000-04-05  
; NUMBER OF SEQ ID NOS: 59  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 7  
; LENGTH: 151  
; TYPE: PRT  
; ORGANISM: Escherichia coli  
US-09-543-407-7

Query Match 66.8%; Score 520; DB 19; Length 151;  
Best Local Similarity 66.9%; Pred. No. 1.1e-46;  
Matches 101; Conservative 21; Mismatches 29; Indels 0; Gaps 0;

Qy 1 MKLLKVAAPAAIVVSGSALAGVVPQWGGGNHNGGSSGPDSTLSIYQYGSANAALALQ 60  
Db 1 MKLLKVAIAAIVFSGSALAGVVPQWGGGNHNGGSSGPDSTLSIYQYGSANAALALQ 60  
Qy 61 SDARKSETTITQSGYNGADVQGADNSTIELTQNGFRNNATIDQWNAKNSDITVGOYGG 120  
Db 61 TDARNSDLTITQHGNGGADVGGSDSSIDLTLQRFGNSATIDQWNGNSEWTVKQFGG 120  
Qy 121 NNAALVNYDQVTRVTHMAHANNATANY 151  
Db 121 GNGAAVDQTASNSSVNVTVQFGNNATAHQY 151

Search completed: August 2, 2004, 15:26:42  
Job time : 168.9 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: August 2, 2004, 14:49:38 ; Search time 17.8 Seconds  
(without alignments)  
888.146 Million cell updates/sec

Title: US-09-543-407-12

Perfect score: 779

Sequence: 1 MKLLKVAAPAAIIVSGSALA.....VTRVTHEMAHANNATANY 151

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 601315 seqs, 104695340 residues

Total number of hits satisfying chosen parameters: 601315

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

1: Pending Parents AA New.\*

2: /cgn2\_6/prodata/2/paa/PCT\_NEW\_COMB.pep.\*

3: /cgn2\_6/prodata/2/paa/US05\_NEW\_COMB.pep.\*

4: /cgn2\_6/prodata/2/paa/US07\_NEW\_COMB.pep.\*

5: /cgn2\_6/prodata/2/paa/US08\_NEW\_COMB.pep.\*

6: /cgn2\_6/prodata/2/paa/US09\_NEW\_COMB.pep.\*

7: /cgn2\_6/prodata/2/paa/US10\_NEW\_COMB.pep.\*

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed,

and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	517	66.4	151	5	US-09-741-873C-4
2	439	56.4	131	5	US-09-741-873C-2
3	97.5	12.5	258	6	US-10-425-115-300390
4	97.5	12.5	295	6	US-10-425-115-312468
5	94.5	12.1	719	5	US-09-248-796A-17559
6	92	11.8	299	6	US-10-170-205B-35751
7	92	11.8	1203	6	US-10-170-205B-741
8	92	11.8	1327	1	PCT-US04-02338-49
9	91.5	11.7	256	6	US-10-425-115-301334
10	91.5	11.7	443	6	US-10-100-683-7608
11	91.5	11.7	841	7	US-60-565-632-7906
12	91.5	11.7	841	7	US-60-579-062-7906
13	90.5	11.6	234	6	US-10-767-701-45603
14	90	11.6	586	1	PCT-US03-24982A-317
15	88	11.3	573	7	US-60-585-632-7907
16	88	11.3	573	7	US-60-579-062-7907
17	87.5	11.2	179	6	US-10-425-115-346132
18	87.5	11.2	892	5	US-09-952-267B-5
19	87.5	11.2	892	6	US-10-872-768-5
20	87.5	11.2	892	6	US-10-872-769-5
21	86	11.0	412	7	US-60-565-632-7905
22	86	11.0	412	7	US-60-579-062-7905
23	86	11.0	520	6	US-10-479-638-21
24	85	10.9	956	6	US-10-093-037A-63
25	84.5	10.8	251	6	US-10-854-439-256
26	83.5	10.7	251	6	US-10-854-439-249

27	82.5	10.6	873	5	US-09-952-267B-13	Sequence 13, Appl
28	82.5	10.6	873	6	US-10-872-768-13	Sequence 13, Appl
29	82.5	10.6	873	6	US-10-872-769-13	Sequence 13, Appl
30	82	10.5	193	6	US-10-425-115-254240	Sequence 254240,
31	82	10.5	386	6	US-10-100-683-10326	Sequence 10326, A
32	82	10.5	386	6	US-10-100-683-10327	Sequence 10327, A
33	82	10.5	386	6	US-10-798-512-100	Sequence 100, App
34	82	10.5	555	1	PCT-US04-13211-3	Sequence 3, Appli
35	82	10.5	555	6	US-10-835-304-3	Sequence 3, Appli
36	82	10.5	1627	6	US-10-170-205E-16659	Sequence 16659, A
37	82	10.5	1905	1	PCT-US04-09388-9	Sequence 9, Appli
38	81.5	10.5	376	6	US-10-491-733-2	Sequence 2, Appli
39	81.5	10.5	508	6	US-10-425-115-285216	Sequence 285216,
40	81.5	10.5	532	6	US-10-425-115-285214	Sequence 285214,
41	81.5	10.5	1871	1	PCT-US03-02038-26	Sequence 26, Appl
42	81.5	10.5	4249	1	PCT-US04-04300-4	Sequence 4, Appli
43	81	10.4	131	7	US-60-565-632-11109	Sequence 11109, A
44	81	10.4	131	7	US-60-579-062-11109	Sequence 11109, A
45	81	10.4	201	6	US-10-425-115-309662	Sequence 309662,

ALIGNMENTS

RESULT 1  
US-09-741-873C-4  
; Sequence 4, Application US/09741873C  
; GENERAL INFORMATION: Staffan  
; APPLICANT: Normark, Staffan  
; APPLICANT: Olsen, Arne  
; TITLE OF INVENTION: Fibronectin Binding Protein As Well As Its Preparation  
; FILE REFERENCE: 012889-084  
; CURRENT APPLICATION NUMBER: US/09/741,873C  
; CURRENT FILING DATE: 2000-12-22  
; PRIOR APPLICATION NUMBER: SE 8801723-1  
; PRIOR FILING DATE: 1998-05-06  
; PRIOR APPLICATION NUMBER: US 08/978,878  
; PRIOR FILING DATE: 1997-11-26  
; PRIOR APPLICATION NUMBER: US 07/347,189  
; PRIOR FILING DATE: 1989-05-04  
; PRIOR APPLICATION NUMBER: US 07/789,437  
; PRIOR FILING DATE: 1991-11-06  
; PRIOR APPLICATION NUMBER: US 07/970,846  
; PRIOR FILING DATE: 1992-11-03  
; PRIOR APPLICATION NUMBER: US 08/187,865  
; PRIOR FILING DATE: 1994-01-28  
; PRIOR APPLICATION NUMBER: US 08/318,519  
; NUMBER OF SEQ ID NOS: 11  
; SOFTWARE: Patent in version 3.0  
; SEQ ID NO 4  
; LENGTH: 151  
; TYPE: PRT  
; ORGANISM: Escherichia coli  
US-09-741-873C-4

Query Match 66.4%; Score 517; DB 5; Length 151;  
Best Local Similarity 66.2%; Pred. No. 1.1e-36;  
Matches 100; Conservative 22; Mismatches 29; Indels 0; Gaps 0;

QY	1	MKLLKVAFAAIVTWSGSALAGVVPQWGGGNNHGGNSGPDSTLSIYQYGSANAALQ	60
Db	1	MKLLKVAIAAIVFSGSAVAGVVPYGGGNNHGGNSGPNSELNIYQYGGNSALALQ	60
QY	61	SDARKSETITQSGYNGADVGQGDADNSTIELTQNGFNNATIDOWNAKNSDITVQYGG	120
Db	61	TDARNSDLTITQHGCGNGADVGQGDSSIDLITQRFNGSATLDQNGKNSMTYKQFGG	120
QY	121	NNAALVNYDQLVTRVVVTHEMAHANNATANY	151
Db	121	NGRAAVDQTASNSVNVTVQVGFNNATAHQY	151

## RESULT 2

US-09-741-873C-2  
; Sequence 2, Application US/09741873C  
; GENERAL INFORMATION:  
; APPLICANT: Normark, Staffan  
; APPLICANT: Olsen, Arne  
; TITLE OF INVENTION: Fibronectin Binding Protein As Well As Its Preparation  
; FILE REFERENCE: 012889-084  
; CURRENT APPLICATION NUMBER: US/09/741,873C  
; CURRENT FILING DATE: 2000-12-22  
; PRIOR APPLICATION NUMBER: SE 8801723-1  
; PRIOR FILING DATE: 1998-05-06  
; PRIOR APPLICATION NUMBER: US 08/978,878  
; PRIOR FILING DATE: 1997-11-26  
; PRIOR APPLICATION NUMBER: US 07/347,189  
; PRIOR FILING DATE: 1989-05-04  
; PRIOR APPLICATION NUMBER: US 07/789,437  
; PRIOR FILING DATE: 1991-11-06  
; PRIOR APPLICATION NUMBER: US 07/970,846  
; PRIOR FILING DATE: 1992-11-03  
; PRIOR APPLICATION NUMBER: US 08/187,865  
; PRIOR FILING DATE: 1994-01-28  
; PRIOR APPLICATION NUMBER: US 08/318,519  
; PRIOR FILING DATE: 1994-10-05  
; NUMBER OF SEQ ID NOS: 11  
; SOFTWARE: PatentIn version 3.0  
; SEQ ID NO 2  
; LENGTH: 131  
; TYPE: PRT  
; ORGANISM: Escherichia coli  
US-09-741-873C-2

Query Match 56.4%; Score 439; DB 5; Length 131;  
Best Local Similarity 62.6%; Pred. No. 3.9e-30;  
Matches 82; Conservative 21; Mismatches 28; Indels 0; Gaps 0;  
  
QY 21 GVVPQWGGGNGGNSGPDSTLSIYQVGSANAALQSDARKSETTITQSGYNGAD 80  
DB 1 GVVPQYGGGNGGNSGNSPSELNTYQVGGNSALALQTDARNSDLITQHGNGAD 60  
  
QY 81 VGGADNSTIETQNGFRNATIDOWNAKNSDITVQYGGNNAALVNYDQVTRVTHM 140  
DB 61 VGGSDSSIDLQRFNGSATLDQNGKSEMTVQFGGNGAAVDQTASNSVNVTVQ 120  
  
QY 141 AHANNATANCY 151  
DB 121 GFGNNATAHQY 131

## RESULT 3

US-10-425-115-300390  
; Sequence 300390, Application US/10425115  
; GENERAL INFORMATION:  
; APPLICANT: La Rosa, Thomas J.  
; APPLICANT: Kovalic, David K.  
; APPLICANT: Zhou, Yihua  
; APPLICANT: Cao, Yongwei  
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With  
; FILE REFERENCE: 38-21(53222)B  
; CURRENT APPLICATION NUMBER: US/10/425,115  
; CURRENT FILING DATE: 2003-04-28  
; NUMBER OF SEQ ID NOS: 369326  
; SEQ ID NO 300390  
; LENGTH: 258  
; TYPE: PRT  
; ORGANISM: Zea mays  
; FEATURE:  
; NAME/KEY: unsure  
; LOCATION: (1)..(258)  
; OTHER INFORMATION: unsure at all xaa locations  
; FEATURE:  
; OTHER INFORMATION: Clone ID: MRT4577\_37025C.1.pep

## US-10-425-115-300390

Query Match 12.5%; Score 97.5; DB 6; Length 258;  
Best Local Similarity 26.6%; Pred. No. 0.86;  
Matches 34; Conservative 18; Mismatches 53; Indels 23; Gaps 5;  
  
QY 17 SALAGVPMQGGGNGGNSGPDSTLSIYQVGSANAALQSDARKSETTITQSG-- 74  
DB 78 SSIAG-----GGGGGCGGGTNGSGSGSGSYGSGSSTAASGSPSGNYADAEGKAG 132  
  
QY 75 -----YNGCA--DVGGADNSTIEL--TNGFRNATIDOWNAKNSDITVQYGG 120  
DB 133 GGMGGGANGAYGSGAGGCGVGVVALAPSGNYNGAADATCGGSG--AGGGHGG 190  
  
QY 121 NNAALVNY 128  
DB 191 GAAGAPSY 198

## RESULT 4

US-10-425-115-312468  
; Sequence 312468, Application US/10425115  
; GENERAL INFORMATION:  
; APPLICANT: La Rosa, Thomas J.  
; APPLICANT: Kovalic, David K.  
; APPLICANT: Zhou, Yihua  
; APPLICANT: Cao, Yongwei  
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With  
; FILE REFERENCE: 38-21(53222)B  
; CURRENT APPLICATION NUMBER: US/10/425,115  
; CURRENT FILING DATE: 2003-04-28  
; NUMBER OF SEQ ID NOS: 369326  
; SEQ ID NO 312468  
; LENGTH: 295  
; TYPE: PRT  
; ORGANISM: Zea mays  
; FEATURE:  
; OTHER INFORMATION: Clone ID: MRT4577\_48027C.1.pep  
US-10-425-115-312468

Query Match 12.5%; Score 97.5; DB 6; Length 295;  
Best Local Similarity 23.7%; Pred. No. 1;  
Matches 38; Conservative 15; Mismatches 50; Indels 25; Gaps 4;  
  
QY 26 WGGGNGGNSGNSGPDSTLSIYQVGSANAALQSDARKSETTITQSGYNGADVQGA 85  
DB 138 YGGGYSGGGYSYG-GYAANGYGVGSGSGNYSNAGGYSGS-----DGYNGAASGGA 192  
  
QY 86 DNSTIELTQNGFRN-----NATIDOWNAKN-----SDITVQYGGNNAAL 125  
DB 193 NNLSSGYNSNGRYNTIGSSDGTGYSNPYNGAGNYNTGGSSSGTGLGFGGGGFGG 252  
  
QY 126 VNYDQVLT 133  
DB 253 NNFAGNVT 260

## RESULT 5

US-09-248-796A-17559  
; Sequence 17559, Application US/09248796A  
; GENERAL INFORMATION:  
; APPLICANT: Keith Weinstein et al  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO CANDIDA ALBICANS  
; TITLE OF INVENTION: FOR DIAGNOSTICS AND THERAPEUTICS  
; FILE REFERENCE: 107196.132  
; CURRENT APPLICATION NUMBER: US/09/248,796A  
; CURRENT FILING DATE: 1999-02-12  
; PRIOR APPLICATION NUMBER: US 60/074,725  
; PRIOR FILING DATE: 1998-02-13  
; PRIOR APPLICATION NUMBER: US 60/096,409  
; PRIOR FILING DATE: 1998-08-13  
; NUMBER OF SEQ ID NOS: 28208

```

; SEQ ID NO 17559
; LENGTH: 719
; TYPE: PRT
; ORGANISM: Canis
US-09-248-796A-175

```

Query Match	12.1%;	Score 94.5;	DB 5;	Length 719;
Best Local Similarity	25.8%;	Pred. No. 4.9;		
Matches 31;	Conservative	24;	Mismatches 42;	Indels 23;
				Gaps 4;

```

RESULT 6
US-10-170-205E-35751
; Sequence 35751, Application US/10170205E
; GENERAL INFORMATION:
; APPLICANT: ADAMS, Mark
; ; TITLE OF INVENTION: DEVICES, SUCH AS ARRAYS, COMPRISED OF HUMAN PROTEINS OR PROTEIN
; ; TITLE OF INVENTION: CAPTURE AGENTS, AND USES THEREOF
; FILE REFERENCE: CL001361
; CURRENT APPLICATION NUMBER: US/10/170,205E
; CURRENT FILING DATE: 2002-06-13
; NUMBER OF SEQ ID NOS: 40312
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 35751
; LENGTH: 299
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-170-205E-35751

```

```

RESULT 7
US-10-170-205E-741
; Sequence 741, Application US/10170205E
; GENERAL INFORMATION:
; APPLICANT: ADAMS, Mark
; TITLE OF INVENTION: DEVICES, SUCH AS ARRAYS, COMPRISED OF HUMAN PROTEINS OR PROTEIN
; TITLE OF INVENTION: CAPTURE AGENTS, AND USES THEREOF
; FILE REFERENCE: CL001381
; CURRENT APPLICATION NUMBER: US/10/170,205E
; CURRENT FILING DATE: 2002-06-13
; NUMBER OF SEQ ID NOS: 40312
; SOFTWARE: Patentin version 3.2
; SEQ ID NO 741
; LENGTH: 1203
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-170-205E-741

```

Db	99	VAAAPWPAVSTSAAGVAPNPAGSGNPSSSSPTSS-SSSSPSPGCSLAESPEAA	157
Qy	65	KSEITIT-----QSGVNGADVQCADNSTIELTQNG---FRNNATIOWNAKSDI	113
Db	158	GYSVSTAPLPGCAAGPGTGVPAVSGALRELLEACRNGDVSRVKELVDAAANNKDM	212

```

RESULT 8
PCT-US04-02338-49
; Sequence 49, Application PC/TUS0402338
; GENERAL INFORMATION:
; APPLICANT: EXELIXIS, INC.
; TITLE OF INVENTION: MAPCAXS AS MODIFIERS OF THE APC AND AXIN PATHWAYS AND METHODS OF
; TITLE OF INVENTION: USE
; FILE REFERENCE: EX04-003C-PC
; CURRENT APPLICATION NUMBER: PCT/US04/02338
; CURRENT FILING DATE: 2004-01-28
; PRIOR APPLICATION NUMBER: US60/443,484
; PRIOR FILING DATE: 2003-01-29
; PRIOR APPLICATION NUMBER: US60/447,358
; PRIOR FILING DATE: 2003-02-11
; PRIOR APPLICATION NUMBER: US60/461,789
; PRIOR FILING DATE: 2003-04-10
; PRIOR APPLICATION NUMBER: US60/470,684
; PRIOR FILING DATE: 2003-05-14
; PRIOR APPLICATION NUMBER: US60/479,650
; PRIOR FILING DATE: 2003-06-19
; NUMBER OF SEQ ID NOS: 54
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 49
; LENGTH: 1327
; TYPE: prt
; ORGANISM: Homo sapiens
PCT-US04-02338-49

```

```

RESULT 9
US-10-425-115-301334
; Sequence 301334, Application US/10425115
; GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J.
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
; TITLE OF INVENTION: Plants
; FILE REFERENCE: 38-21(53222)B
; CURRENT APPLICATION NUMBER: US/10/425,115
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 369326
; SEQ ID NO 301334
; LENGTH: 256
; TYPE: PAT
; ORGANISM: Zea mays
; FEATURE:
; OTHER INFORMATION: Clone ID: MET4577_37894C.1.pep
US-10-425-115-301334

```

Query Match	11.7%	Score 91.5;	DB 6;	Length 256;
Best Local Similarity	26.6%	Pred. No. 2.8;		

Matches 33; Conservative 16; Mismatches 52; Indels 23; Gaps 5;  
QY 17 SALAGVVPQGGGNGHNGGSSGPPDSTLSIYQYGSANAALALQSDA-----RKSETTITQ 72  
Db 79 SSVAG-----GGGGGGGGGTNGSGSGSGSGSGTSTAASGSSGNYANAEGKGAG 133  
QY 73 SGYGVNGAD-----VQGGADNSTIEL--TQNGFRNATIDQWNAKNSDITVQYGG 120  
Db 134 GNGGGGADGAYGSGAGGVCQGGESGVALPSSDGYNGGAADATGGGSG--AGGGHG 191  
QY 121 NNA 124  
Db 192 GAGA 195

RESULT 10  
US-10-100-683-7608  
; Sequence 7608, Application US/10100683  
; GENERAL INFORMATION:  
; APPLICANT: Rosen, et al.  
; TITLE OF INVENTION: Human Secreted Proteins  
; FILE REFERENCE: PS900  
; CURRENT APPLICATION NUMBER: US/10/100,683  
; CURRENT FILING DATE: 2002-03-19  
; PRIOR APPLICATION NUMBER: US 60/040,162  
; PRIOR FILING DATE: 1997-03-07  
; PRIOR APPLICATION NUMBER: US 60/043,576  
; PRIOR FILING DATE: 1997-04-11  
; PRIOR APPLICATION NUMBER: US 60/047,601  
; PRIOR FILING DATE: 1997-05-23  
; PRIOR APPLICATION NUMBER: US 60/056,845  
; PRIOR FILING DATE: 1997-08-22  
; PRIOR APPLICATION NUMBER: US 60/043,580  
; PRIOR FILING DATE: 1997-04-11  
; PRIOR APPLICATION NUMBER: US 60/047,599  
; PRIOR FILING DATE: 1997-05-23  
; PRIOR APPLICATION NUMBER: US 60/056,664  
; PRIOR FILING DATE: 1997-08-22  
; PRIOR APPLICATION NUMBER: US 60/043,314  
; PRIOR FILING DATE: 1997-04-11  
; PRIOR APPLICATION NUMBER: US 60/047,632  
; PRIOR FILING DATE: 1997-05-23  
; PRIOR APPLICATION NUMBER: US 60/056,892  
; PRIOR FILING DATE: 1997-08-22  
; Remaining Prior Application data removed - See File Wrapper or PALM.  
; NUMBER OF SEQ ID NOS: 13468  
; SOFTWARE: Patent in Ver. 2.0  
; SEQ ID NO 7608  
; LENGTH: 443  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-10-100-683-7608

Query Match 11.7%; Score 91.5; DB 6; Length 443;  
Best Local Similarity 26.0%; Pred. No. 5.1;  
Matches 33; Conservative 18; Mismatches 43; Indels 33; Gaps 6;  
QY 21 GVVPQ---WCGGNGHNGGSSGPPDSTLSIYQ--YGSANAALALQSDKSETTITQSGV 75  
Db 185 GMPQAGAPWQGG--NGGPPNFGTNGAVAPQYGVRA-----SNQEGCTNPPPSGS 237  
QY 76 GNGAD---VQGGADNSTIELTQNGFRNATI-----DQWNAKNSDIT 114  
Db 238 GGGSSNGGSGSGSGSGSGGNDNNGSSGSGSGSGSGSGSGSGSGSGSGSGSGSGSGSG 297  
QY 115 VQYVGN 121  
Db 298 SGNHGS 304

RESULT 11  
US-60-565-632-7906  
; Sequence 7906, Application US/60565632

GENERAL INFORMATION:  
; APPLICANT: Monsanto Technology, LLC  
; APPLICANT: Baum, James A  
; APPLICANT: Kovalic, David K.  
; APPLICANT: Larosa, Thomas J.  
; APPLICANT: Lu, Maolong  
; APPLICANT: Muryikwa, Tichifa R. I.  
; APPLICANT: Roberts, James K.  
; APPLICANT: Wu, Wei  
; APPLICANT: Zhang, Bei  
; TITLE OF INVENTION: Methods for Genetic Control of Insect Infestations in Plants and  
; TITLE OF INVENTION: Compositions Thereof  
; FILE REFERENCE: 38-21 (53403)B  
; CURRENT APPLICATION NUMBER: US/60/565,632  
; CURRENT FILING DATE: 2004-04-27  
; NUMBER OF SEQ ID NOS: 15449  
; SOFTWARE: Patent in version 3.2  
; SEQ ID NO 7906  
; LENGTH: 841  
; TYPE: PRT  
; ORGANISM: Diabrotica virgifera  
; FEATURE:  
; NAME/KEY: misc feature  
; LOCATION: (810)..(810)  
; OTHER INFORMATION: Xaa can be any naturally occurring amino acid  
US-60-565-632-7906  
Query Match 11.7%; Score 91.5; DB 7; Length 841;  
Best Local Similarity 32.6%; Pred. No. 11;  
Matches 44; Conservative 12; Mismatches 48; Indels 31; Gaps 9;  
QY 30 GNHGGG--NSSGPDSTLSIYQGSANAALA-----LQSDKSETTITQSGYNGAD 80  
Db 269 GNEGTGAENNAADATQDAAQ--GSTNEAENNAADVQDAAQENAGAENSAD 327  
QY 81 VQGGADN-STIELTQN-----GFENNATIDQWNAKNSDITVQYGGNNAALVNYDQLVTR 134  
Db 328 AAQGTGNGAENGTGNADPAGNDNGAA-----AENSGNENGTAAGNA---NPD----- 374  
QY 135 VYTHEMAHAN-NATA 148  
Db 375 -VQDAAQVNDNGTA 388

RESULT 12  
US-60-579-062-7906  
; Sequence 7906, Application US/60579062  
; GENERAL INFORMATION:  
; APPLICANT: Baum, James A  
; APPLICANT: Kovalic, David K  
; APPLICANT: Larosa, Thomas J  
; APPLICANT: Lu, Maolong  
; APPLICANT: Muryikwa, Tichifa R. I.  
; APPLICANT: Roberts, James K.  
; APPLICANT: Wu, Wei  
; APPLICANT: Zhang, Bei  
; TITLE OF INVENTION: Methods for Genetic Control of Insect Infestations in Plants and  
; TITLE OF INVENTION: Compositions thereof  
; FILE REFERENCE: 38-21 (53403) C  
; CURRENT APPLICATION NUMBER: US/60/579,062  
; CURRENT FILING DATE: 2004-06-11  
; NUMBER OF SEQ ID NOS: 41445  
; SOFTWARE: Patent in version 3.2  
; SEQ ID NO 7906  
; LENGTH: 841  
; TYPE: PRT  
; ORGANISM: Diabrotica virgifera  
; FEATURE:  
; NAME/KEY: misc feature  
; LOCATION: (810)..(810)  
; OTHER INFORMATION: Xaa can be any naturally occurring amino acid  
US-60-579-062-7906



Query Match 11.7%; Score 91.5; DB 7; Length 841;  
Best Local Similarity 32.6%; Pred. No. 11;  
Matches 44; Conservative 12; Mismatches 48; Indels 31; Gaps 9;  
QY 30 GNHNGGG--NSSGPDSTLSIYQGSANAALA-----LQSDARKSETTITQSYGNGAD 80  
Db 269 GNENGTGAENANADQTDAAQ--GSTNEAENANADVQNDAAQENGAENGAENSGNAD 327  
QY 81 VQGGADN--STIELTON-----GFRNATIDQWNAKNSDITVQYGGNNAALVNYDOLVTR 134  
Db 328 AAQGTGNGAAENTGNADPAQNDNGAA-----AENSGNENGTAAGNNA---NPD----- 374  
QY 135 VVTHEMAHAN--NATA 148  
Db 375 -VONDAQVNDNGTA 388  
RESULT 13  
US-10-767-701-45603  
; Sequence 45603, Application US/10767701  
; GENERAL INFORMATION:  
; APPLICANT: Kovalic, David K.  
; APPLICANT: Zhou, Yihua  
; APPLICANT: Cao, Yongwei  
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With  
; FILE REFERENCE: 38-21(53535)B  
; CURRENT APPLICATION NUMBER: US/10767, 701  
; CURRENT FILING DATE: 2004-01-29  
; NUMBER OF SEQ ID NOS: 63128  
; SEQ ID NO 45603  
; LENGTH: 234  
; TYPE: PRT  
; ORGANISM: Sorghum bicolor  
; FEATURE:  
; NAME/KEY: unsure  
; LOCATION: (1)..(234)  
; OTHER INFORMATION: unsure at all xaa locations  
; FEATURE:  
; OTHER INFORMATION: Clone ID: SORBI-28MAY03-C82834\_1.pep  
US-10-767-701-45603  
Query Match 11.6%; Score 90.5; DB 6; Length 234;  
Best Local Similarity 24.2%; Pred. No. 3;  
Matches 39; Conservative 12; Mismatches 45; Indels 65; Gaps 6;  
QY 1 MKLLKVAFAAIVV--SGSALAGVVPQW---GGGNHNGGN-----SSGPDSTL 45  
Db 1 MATTKAALCFVLLGIGGANAARVARYVAGGGGGGGGGRGWRGGASRWGSGSGC 60  
QY 46 SIY-----QYGSANAALALQSDARKSETTITQSYGNGADVG 82  
Db 61 GXYGEAGSGXAYAQGGGGGGGGGGYGGSG-----SGYGGSGGYG 102  
QY 83 QGADNSTIELTONGFRNATIDQWNAKNSDITVQYGGNNA 123  
Db 103 QAGSGS-----NGGAYAQGAQGGGGGGGGYGGSGS 134  
RESULT 14  
PCT-US03-24982A-317  
; Sequence 317, Application PC/TUS0324982A  
; GENERAL INFORMATION:  
; APPLICANT: Syngenta Participations AG  
; APPLICANT: Stam, Lynn  
; APPLICANT: Kamdar, Kim  
; APPLICANT: Spana, Eric  
; APPLICANT: Bachmann, Jane  
; TITLE OF INVENTION: Nucleic Acid Sequences from Drosophila Melanogaster that Encode P  
; FILE REFERENCE: 70131WOPCT  
; CURRENT APPLICATION NUMBER: PCT/US03/24982A  
; CURRENT FILING DATE: 2003-08-08

; PRIOR APPLICATION NUMBER: 60/422,377  
; PRIOR FILING DATE: 2002-10-29  
; NUMBER OF SEQ ID NOS: 381  
; SOFTWARE: PatentIn version 3.1  
; SEQ ID NO 317  
; LENGTH: 586  
; TYPE: PRT  
; ORGANISM: Drosophila melanogaster  
PCT-US03-24982A-317  
Query Match 11.6%; Score 90; DB 1; Length 586;  
Best Local Similarity 25.5%; Pred. No. 9.5;  
Matches 28; Conservative 12; Mismatches 44; Indels 26; Gaps 3;  
QY 27 GG--GGNHNGG-----NSSGPDSTLSIYQGSANAALALQSDARKSETTITQSYGNG 78  
Db 347 GGNFGNNNGGGFGNSGNNFNSPGVNNFGNNGSGGFGNGGNGFNVSNGVNF 406  
QY 79 ADVGQADNSTIELTONGFRNATIDQWNAKNSDITVQYGGNNAALVNY 128  
Db 407 GPIGGGRNNNGNFGNSGF-----GNFGNNNVGSNF 438  
RESULT 15  
US-60-565-632-7907  
; Sequence 7907, Application US/60565632  
; GENERAL INFORMATION:  
; APPLICANT: Monsanto Technology, LLC  
; APPLICANT: Baum, James A  
; APPLICANT: Kovalic, David K.  
; APPLICANT: Larosa, Thomas J.  
; APPLICANT: Lu, Maolong  
; APPLICANT: Munyikwa, Tichifa R. I.  
; APPLICANT: Roberts, James K.  
; APPLICANT: Wu, Wei  
; APPLICANT: Zhang, Bei  
; TITLE OF INVENTION: Methods for Genetic Control of Insect Infestations in Plants and  
; FILE REFERENCE: 38-21(53403)B  
; CURRENT APPLICATION NUMBER: US/60/565,632  
; CURRENT FILING DATE: 2004-04-27  
; NUMBER OF SEQ ID NOS: 15449  
; SOFTWARE: PatentIn version 3.2  
; SEQ ID NO 7907  
; LENGTH: 573  
; TYPE: PRT  
; ORGANISM: Diabrotica virgifera  
US-60-565-632-7907  
Query Match 11.3%; Score 88; DB 7; Length 573;  
Best Local Similarity 26.2%; Pred. No. 14;  
Matches 38; Conservative 16; Mismatches 49; Indels 42; Gaps 7;  
QY 30 GNHNGGG--NSSGPDSTLSIYQGSANAALA-----LQSDARKSETTITQSYGNGAD 80  
Db 241 GNENGTGAENANADQTDVAQ--GSTNEAENANADVQNDAAQENGAENGAENSGNAD 299  
QY 81 VQGGADNSTI-----ELTQNGFRNATIDQWNAKNSDITVQYGGNNA 124  
Db 300 AAQGTGNGAAENTGNADPAQNDNGAAENSGNENGTAAGNNA--NADVQ-----NDAA 352  
QY 125 LVNYDQLVTRVVTHEMAHANATAN 149  
Db 353 QVN-----DNGAAENNGNAD 368  
Search completed: August 2, 2004, 15:29:50  
Job time : 17.8 secs





Db 121 NNAALVNTASDSSVMVRQVGFGNATANQY 151

RESULT 2

AI0635

major curlin chain precursor [imported] - Salmonella enterica subsp. enterica serovar Typhimurium

C:Species: Salmonella enterica subsp. enterica serovar Typhi

A:Note: this species has also been called Salmonella typhimurium

C:Date: 09-Nov-2001 #sequence\_revision 09-Nov-2001 #text\_change 18-Nov-2002

C:Accession: AI0635

R:Parkhill, J.; Dougan, G.; James, K.D.; Thomson, N.R.; Pickard, D.; Wain, J.; Churcher, T.; Connerton, P.; Cronin, A.; Davis, P.; Davies, R.M.; Dowd, L.; White, N.; Farrar, S.; Moule, S.; O'Gaora, P.

A:Title: The complete genome sequence of a multiple drug resistant Salmonella enterica serovar Typhimurium

A:Reference number: AB0502; MUID:21534947; PMID:11677608

A:Accession: AI0635

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-151 <PAR>

A:Cross-references: GB:AL513382; PIDN:CAD08268.1; PID:g16502315; GSPDB:GN00176

C:Genetics:

A:Gene: STY1181

Query Match 89.2%; Score 695; DB 2; Length 151;

Best Local Similarity 90.1%; Pred. No. 1.8e-50;

Matches 136; Conservative 4; Mismatches 11; Indels 0; Gaps 0;

QY 1 MKLLKVAFAAIVVSGSALAGVVPQWGGGNGHNGSGSPDSTLSIYQVGSANAALALQ 60

Db 1 MKLLKVAFAAIVVSGSALAGVVPQWGGGNGHNGSGSPDSTLSIYQVGSANAALALQ 60

QY 61 SDARKSETTITQSGYNGADVGQADNSTIELTQNGFRNNATIDQWNAKNSDITVQYGG 120

Db 61 SDARKSETTITQSGYNGADVGQADNSTIELTQNGFRNNATIDQWNAKNSDITVQYGG 120

QY 121 NNAALVNDQLVTRVVTTHMAHANNTANQY 151

Db 121 NNAALVNTASDSSVMVRQVGFGNATANQY 151

RESULT 3

S70788

curlin protein csgA precursor - Escherichia coli (strain K-12)

N:Alternate names: csgA protein; major curlin protein

C:Species: Escherichia coli

C:Date: 12-Feb-1998 #sequence\_revision 20-Feb-1998 #text\_change 01-Mar-2002

C:Accession: S70788; G64846; S31202; S34560; S34559

R:Hammar, M.; Arngvist, A.; Bian, Z.; Olsen, A.; Normark, S.

Mol. Microbiol. 18, 661-670, 1995

A:Title: Expression of two csg operons is required for production of fibronectin- and curli

A:Reference number: S70783; MUID:96414468; PMID:8817489

A:Accession: S70788

A:Status: nucleic acid sequence not shown; translation not shown

A:Molecule type: DNA

A:Residues: 1-151 <HAM>

A:Cross-references: EMBL:X90754; NID:g1147558; PIDN:CAA62282.1; PID:g1147564

A:Experimental source: strain K12, substrain W3110

A:Note: the nucleotide sequence was submitted to the EMBL Data Library, August 1995

R:Blattner, F.R.; Plunkett III, G.; Bloch, C.A.; Perna, N.T.; Burland, V.; Riley, M.; Cohen, D.; Rose, D.J.; Mau, B.; Shaoh, Y.

A:Title: The complete genome sequence of Escherichia coli K-12.

A:Reference number: A64720; MUID:97426617; PMID:9278503

A:Accession: G64846

A:Status: nucleic acid sequence not shown; translation not shown

A:Molecule type: DNA

A:Residues: 1-151 <BLAT>

A:Cross-references: GB:AB000205; GB:U00096; NID:g1787265; PIDN:AAC74126.1; PID:g1787279;

A:Experimental source: strain K-12, substrain M61655

R:Olsen, A.; Arngvist, A.; Hammar, M.; Sukupolvi, S.; Normark, S.

Mol. Microbiol. 7, 523-536, 1993

Query Match 66.8%; Score 520; DB 2; Length 151;

Best Local Similarity 66.9%; Pred. No. 4.7e-36;

Matches 101; Conservative 21; Mismatches 29; Indels 0; Gaps 0;

QY 1 MKLLKVAFAAIVVSGSALAGVVPQWGGGNGHNGSGSPDSTLSIYQVGSANAALALQ 60

Db 1 MKLLKVAFAAIVVSGSALAGVVPQWGGGNGHNGSGSPDSTLSIYQVGSANAALALQ 60

QY 61 SDARKSETTITQSGYNGADVGQADNSTIELTQNGFRNNATIDQWNAKNSDITVQYGG 120

Db 61 TDARNSDLTITQHGNGGADVGQSDSSIDLITQGFNSATLTDQWNGKNSMTVKQFGG 120

QY 121 NNAALVNDQLVTRVVTTHMAHANNTANQY 151

Db 121 NGGAADVDTASNSVNTQVGFGNATANQY 151

RESULT 4

D90806

curlin major subunit CsgA [imported] - Escherichia coli (strain O157:H7, substrain RIMD 0806)

C:Species: Escherichia coli

C:Date: 18-Jul-2001 #sequence\_revision 18-Jul-2001 #text\_change 18-Jul-2001

C:Accession: D90806

R:Hayashi, T.; Makino, K.; Ohnishi, M.; Kurokawa, K.; Ishii, K.; Yokoyama, K.; Han, C.G.; Gasawara, N.; Yasunaga, T.; Kuhara, S.; Shiba, T.; Hattori, M.; Shinagawa, H.

DNA Res. 8, 11-22, 2001

A:Title: Complete genome sequence of enterohemorrhagic Escherichia coli O157:H7 and genomic islands

A:Reference number: A39629; MUID:21156231; PMID:11258796

A:Accession: D90806

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-152 <HAY>

A:Cross-references: GB:BA000007; PIDN:BA034843.1; PID:g13360880; GSPDB:GN00154

A:Experimental source: strain O157:H7, substrain RIMD 0509952

C:Genetics:

A:Gene: ECs1420

Query Match 64.4%; Score 501.5; DB 2; Length 152;

Best Local Similarity 65.8%; Pred. No. 1.6e-34;

Matches 100; Conservative 21; Mismatches 30; Indels 1; Gaps 1;

QY 1 MKLLKVAFAAIVVSGSALAGVVPQWGGGNGHNGSGSPDSTLSIYQVGSANAALALQ 59

Db 1 MKLLKVAFAAIVVSGSALAGVVPQWGGGNGHNGSGSPDSTLSIYQVGSANAALALQ 60

QY 60 QSDARKSETTITQSGYNGADVGQADNSTIELTQNGFRNNATIDQWNAKNSDITVQYGG 119

Db 61 QADARNSLDITQHGNGGADVQGGSDSSIDITQRFNGSATLDQWNGKDSHMTVKQF 120  
QY 120 GNNAAALVNYDQVTRVVTTHMANNATANQY 151  
Db 121 GNGAAVDQTAASNTVNTVQVGGNNATAHQY 152  
RESULT 5  
H85665  
hypothetical protein csgA [imported] - Escherichia coli (strain O157:H7, substrain EDL93  
C;Species: Escherichia coli  
C;Date: 16-Feb-2001 #sequence\_revision 16-Feb-2001 #text\_change 14-Sep-2001  
C;Accession: H85665  
R;Perna, N.T.; Plunkett III, G.; Burland, V.; Mau, B.; Glasner, J.D.; Rose, D.J.; Mayhew  
iller, L.; Grotbeck, E.J.; Davis, N.W.; Lim, A.; Dimalanta, E.; Potamousis, K.; Apodaca,  
Nature 409, 529-533, 2001  
A;Title: Genome sequence of enterohemorrhagic Escherichia coli O157:H7.  
A;Reference number: A85480; MUID:21074935; PMID:11206551  
A;Accession: H85665  
A;Status: preliminary  
A;Molecule type: DNA  
A;Residues: 1-152 <STO>  
A;Cross-references: GB:AE0051174; NID:g12514574; PIDN:AA655788.1; GSPDB:GN00145; UMGF:216  
A;Experimental source: strain O157:H7, substrain EDL933  
C;Genetics:  
A;Gene: csgA  
Query Match 64.4%; Score 501.5; DB 2; Length 152;  
Best Local Similarity 65.8%; Pred. No. 1.6e-34;  
Matches 100; Conservative 21; Mismatches 30; Indels 1; Gaps 1;  
QY 1 MKLLKVAAPAAIVVSSGALAGVVPQW-GGGGNGGNGSGGPDSTLSIIYQGSANAALAL 59  
Db 1 MKLLKVAATAIAIVFSGALAGVVPQYGGGGNGGNGSGGPNSELNIIYQGGNGALAL 60  
QY 60 QSDARKSETTITQSGYNGGADVQGGADNSTIELTQGFNRNATIDQWNAKNSDITVQYQ 119  
Db 61 QADARNSLDITQHGNGGADVQGGSDSSIDITQRFNGSATLDQWNGKDSHMTVKQF 120  
QY 120 GNNAAALVNYDQVTRVVTTHMANNATANQY 151  
Db 121 GNGAAVDQTAASNTVNTVQVGGNNATAHQY 152  
RESULT 6  
S42136  
cnjB protein - Tetrahymena thermophila  
C;Species: Tetrahymena thermophila  
C;Date: 19-Mar-1997 #sequence\_revision 18-Jul-1997 #text\_change 07-Dec-1999  
C;Accession: S42136; S42135; S03650  
R;Taylor, F.M.; Martindale, D.W.  
submitted to the EMBL Data Library, October 1992  
A;Reference number: S42136  
A;Accession: S42136  
A;Molecule type: DNA  
A;Residues: 1-1748 <TAY>  
A;Cross-references: EMBL:L03710; NID:g161751; PID:g161752  
R;Taylor, F.M.; Martindale, D.W.  
Nucleic Acids Res 21, 4610-4614, 1993  
A;Title: Retroviral-type zinc fingers and glycine-rich repeats in a protein encoded by c  
A;Reference number: S42135; MUID:94051569; PMID:8233798  
A;Accession: S42135  
A;Molecule type: DNA  
A;Residues: 1164-1174;1179-1198;1233-1252;1285-1293;1316-1326;1331-1341;1343-1  
A;Cross-references: EMBL:L03710  
R;Martindale, D.W.; Taylor, F.M.  
Nucleic Acids Res 15, 2189-2201, 1988  
A;Title: Multiple introns in a conjugation-specific gene from Tetrahymena thermophila.  
A;Reference number: S03650; MUID:88189811; PMID:3357771  
A;Accession: S03650  
A;Molecule type: DNA  
A;Residues: 236-250,'I',252-255,'N',257-273 <MAR>

A;Cross-references: EMBL:X06462  
C;Genetics:  
A;Gene: cnjB  
A;Genetic code: SGC5  
A;Introns: 85/3; 136/1; 157/3; 201/2; 290/2; 327/3; 499/1; 573/2; 607/3; 708/3; 777/3; 88  
C;Keywords: zinc finger  
F;1164-1450/Region: glycine-rich  
F;1451-1464/Region: zinc finger CCHC motif  
F;1478-1491/Region: zinc finger CCHC motif  
F;1501-1514/Region: zinc finger CCHC motif  
F;1530-1543/Region: zinc finger CCHC motif  
F;1555-1568/Region: zinc finger CCHC motif  
F;1579-1592/Region: zinc finger CCHC motif  
F;1602-1615/Region: zinc finger CCHC motif  
F;1628-1748/Region: glycine-rich  
Query Match 13.6%; Score 106; DB 2; Length 1748;  
Best Local Similarity 32.1%; Pred. No. 1;  
Matches 35; Conservative 13; Mismatches 33; Indels 28; Gaps 5;  
QY 25 QWGGGNGG---GNSSGPDSTLSIIYQGSANAALALQSDARKSETTIT---QSGYGN 77  
Db 1640 QFGGGNGSGGSGWTSSGDN-----CQSNVQESTTSSCGWSSSGN 1685  
QY 78 GADVQGGADNSTIELTQGFNRNATIDQWNAKNSDITVQY--YGGNAA 124  
Db 1686 QTGGGNGSDN-----QQQNTGNTGGGNGSSNQTNNSSWGNNQA 1729  
RESULT 7  
E95965  
hypothetical glycine-rich protein [imported] - Sinorhizobium meliloti (strain 1021) magat  
C;Species: Sinorhizobium meliloti  
C;Date: 24-Aug-2001 #sequence\_revision 24-Aug-2001 #text\_change 30-Sep-2001  
C;Accession: E95965  
R;Finan, T.M.; Weidner, S.; Wong, K.; Buhrmester, J.; Chain, P.; Vorholter, F.J.; Hernan  
Proc. Natl. Acad. Sci. U.S.A. 98, 9889-9894, 2001  
A;Title: The complete sequence of the 1,683-kb pSymB megaplasmid from the N2-fixing endo  
A;Reference number: A95842; MUID:21396508; PMID:11481431  
A;Accession: E95965  
A;Status: preliminary  
A;Molecule type: DNA  
A;Residues: 1-2174 <KUR>  
A;Cross-references: GB:AL591985; PIDN:CAC49389.1; PID:g15140875; GSPDB:GN00167  
A;Experimental source: strain 1021, megaplasmid pSymB  
R;Galibert, F.; Finan, T.M.; Long, S.R.; Puhler, A.; Abola, P.; Ampe, F.; Barloy-Hubler,  
pella, D.; Chain, P.; Cowie, A.; Davis, R.W.; Dreano, S.; Federspiel, N.A.; Fisher, R.F.;  
J.; Hyman, R.W.; Jones, T.  
Science 293, 668-672, 2001  
A;Authors: Kahn, D.; Kahn, M.L.; Kalman, S.; Keating, D.H.; Kiss, E.; Komp, C.; Lelaure,  
hebaull, P.; Vandenbol, M.; Vorholter, F.J.; Weidner, S.; Wells, D.H.; Wong, K.; Feh, K.  
A;Title: The composite genome of the legume symbiont Sinorhizobium meliloti.  
A;Reference number: A96039; MUID:21368234; PMID:11474104  
C;Contents: annotation  
C;Genetics:  
A;Gene: Smb21548  
A;Genome: plasmid  
Query Match 13.4%; Score 104; DB 2; Length 2174;  
Best Local Similarity 27.1%; Pred. No. 1.9;  
Matches 35; Conservative 16; Mismatches 44; Indels 34; Gaps 6;  
QY 11 AIWVSGSALAGVVPQ--WGGGNGGNGSGGPDSTLSIIYQGSANAALALQSDARKSET 68  
Db 693 AIATAGAGAVGLAQSIGGGNGN---GGNATGGDAGFGSFOIGGGGGG----- 737  
QY 69 TITQSGYNGADVQ-----QGADNSTI--ELTQNGFRNATIDQWNAK---NSDITV 115  
Db 738 ----GGYANTANVFGKGLTLITQGHAAQIVAQSVGGGGTGGTASSYSAGIGFTASVAV 793  
QY 116 QYGGNNAA 124  
Db 794 GGTGGNGGA 802

RESULT 8	
AH0635	
nucleation component of curlin monomers [imported] - Salmonella enterica subsp. enterica	
C:Species: Salmonella enterica subsp. enterica serovar Typhi	
C>Note: this species has also been called Salmomella typhi	
C>Date: 09-Nov-2001 #sequence_revision 09-Nov-2001 #text_change 18-Nov-2002	
C:Accession: AH0635	
R:Parkhill, J.; Dougan, G.; James, K.D.; Thomson, N.R.; Pickard, D.; Wain, J.; Churcher, T.; Parkhill, J.; Dougan, G.; James, K.D.; Thomson, N.R.; Pickard, D.; Wain, J.; Churcher, S.; Moule, S.; O'Gaora, P.	
Nature 413, 848-852, 2001	
A:Authors: Parry, C.; Quail, M.; Rutherford, K.; Simmonds, M.; Skelton, J.; Stevens, K.	
A>Title: Complete genome sequence of a multiple drug resistant Salmonella enterica serov	
A:Reference number: AB0502; MUID:21534947; PMID:11677608	
A:Accession: AH0635	
A>Status: preliminary	
A:Molecule type: DNA	
A:Residues: 1-151 <PAR>	
A:Cross-references: GB:AL513382; PIDN:CAD08267.1; PID:G16502314; GSPDB:GN00176	
C:Genetics:	
A:Gene: STY1180	
Query Match 13.0%; Score 101.5; DB 2; Length 151;	
Best Local Similarity 30.2%; Pred. No. 0.15;	
Matches 29; Conservative 17; Mismatches 49; Indels 1; Gaps 1;	
QY	47 IYVGSNAAALQSDARKSETTITQSGYNGADVGQADNSTIETQNGFRNNATIDQW 106
Db	49 IGQVGTNSA-RVQEGSKLLSVISQEGNNRAKVQAGNYNFAIEQTGNANDASISQS 107
QY	107 NAKNSDITVGYGGNNAALVNYDQVTRVVTHEMAH 142
Db	108 AYGNSAIIQKSGNKANITQYGTQKTAVVVQKQSH 143
RESULT 9	
JC6040	
fimbriin protein agfB precursor - Salmonella enteritidis	
C:Species: Salmonella enteritidis	
C>Date: 31-Dec-1996 #sequence_revision 31-Dec-1996 #text_change 08-Oct-1999	
C:Accession: JC6040	
R:Collinson, S.K.; Clouthier, S.C.; Doran, J.L.; Bansen, P.A.; Kay, W.W.	
J. Bacteriol. 178, 662-667, 1996	
A>Title: Salmonella enteritidis agfBAC operon encoding thin, aggregative fimbriae.	
A:Reference number: JC6039; MUID:96146512; PMID:8550497	
A:Accession: JC6040	
A:Molecule type: DNA	
A:Residues: 1-151 <COL>	
A:Cross-references: GB:U43280; NID:G1184712; PIDN:AAC43598.1; PID:G1184713	
A:Experimental source: strain 276755-3b	
C:Genetics:	
A:Gene: agfB	
C:Function:	
A:Description: minor component of thin aggregative fimbriae	
A>Note: fimbriae bind to fibronectin, plasminogen, tissue plasminogen activator	
C:Keywords: fimbria	
F:1-21/Domain: signal sequence #status predicted <SIG>	
F:22-151/Product: fimbriin protein agfB #status predicted <MAT>	
Query Match 12.9%; Score 100.5; DB 2; Length 151;	
Best Local Similarity 30.2%; Pred. No. 0.18;	
Matches 29; Conservative 17; Mismatches 49; Indels 1; Gaps 1;	
QY	47 IYVGSNAAALQSDARKSETTITQSGYNGADVGQADNSTIETQNGFRNNATIDQW 106
Db	49 IGQVGTNSA-RVQEGSKLLSVISQEGNNRAKVQAGNYNFAIEQTGNANDASISQS 107
QY	107 NAKNSDITVGYGGNNAALVNYDQVTRVVTHEMAH 142
Db	108 AYGNSAIIQKSGNKANITQYGTQKTAVVVQKQSH 143

RESULT 10	
AS6038	
DNA-binding protein ovo - fruit fly (Drosophila melanogaster)	
C:Species: Drosophila melanogaster	
C>Date: 01-Dec-1995 #sequence_revision 01-Dec-1995 #text_change 21-Jul-2000	
C:Accession: AS6038	
R:Garfinkel, M.D.; Wang, J.; Liang, Y.; Mahowald, A.P.	
Mol. Cell. Biol. 14, 6809-6818, 1994	
A>Title: Multiple products from the shavenbaby-ovo gene region of Drosophila melanogaster	
A:Reference number: AS6038; MUID:95021209; PMID:7935398	
A:Accession: AS6038	
A>Status: preliminary	
A:Molecule type: mRNA	
A:Residues: 1-1028 <GAR>	
A:Cross-references: GB:U11383; NID:G520526; PIDN:AAB60216.1; PID:G520527	
C:Genetics:	
A:Gene: ovo	
A:Cross-references: FlyBase:FBgn0003028	
Query Match 12.6%; Score 98.5; DB 2; Length 1028;	
Best Local Similarity 24.5%; Pred. No. 2.3;	
Matches 46; Conservative 15; Mismatches 62; Indels 65; Gaps 8;	
QY	3 LLKVAAPAAIVVSGSALAGVVPQWGGGNGGNGGNSGPDSTLSIYQYGSANAALALQSD 62
Db	59 LQNAARAAIYIVSAGS-----GGCTGNGGGAGSPGGSPANSGGGGGG----- 104
QY	63 ARKSETTITQSGYNGADVGQADNSTIETQNGFRNNATIDQNAKNSDI----- 113
Db	105 -----GGNGYINCGVG-GPNN--LDGNLLNFASVSNYNESNFKPHNHHHHQH 152
QY	114 -----TVGQ-----YGGNNAAL-----VNVQ-----LVTRVVTHEMAH 143
Db	153 NNNNNNGGQTSMMGHFPYGNPSAYGILKDEPDIEDEAKIDIGTFAQNIQTWGS 212
QY	144 NNATANQY 151
Db	213 GQFNASAY 220
RESULT 11	
SI6356	
ovo protein - fruit fly (Drosophila melanogaster)	
C:Species: Drosophila melanogaster	
C>Date: 13-Jan-1995 #sequence_revision 13-Jan-1995 #text_change 16-Feb-1997	
C:Accession: SI6356	
R:Mevel-Ninio, M.; Terracol, R.; Kafatos, F.C.	
EMBO J. 10, 2259-2266, 1991	
A>Title: The ovo gene of Drosophila encodes a zinc finger protein required for female ge	
A:Reference number: SI6356; MUID:91293102; PMID:1712294	
A:Accession: SI6356	
A>Status: preliminary	
A:Molecule type: DNA	
A:Residues: 1-1213 <MEV>	
A:Cross-references: EMBL:X59772	
C:Genetics:	
A:Gene: FlyBase:ovo	
A:Cross-references: FlyBase:FBgn0003028	
A:Introns: 931/3; 1152/3	
Query Match 12.6%; Score 98.5; DB 2; Length 1213;	
Best Local Similarity 24.5%; Pred. No. 2.8;	
Matches 46; Conservative 15; Mismatches 62; Indels 65; Gaps 8;	
QY	3 LLKVAAPAAIVVSGSALAGVVPQWGGGNGGNGGNSGPDSTLSIYQYGSANAALALQSD 62
Db	422 LQNAARAAIYIVSAGS-----GGCTGNGGGAGSPGGSPANSGGGGGG----- 467
QY	63 ARKSETTITQSGYNGADVGQADNSTIETQNGFRNNATIDQNAKNSDI----- 113
Db	468 -----GGNGYINCGVG-GPNN--LDGNLLNFASVSNYNESNFKPHNHHHHQH 515

QY 114 -----TVGQ-----YGNNAAL-----VNYDQ-----LVTRVVTHEMAHA 143  
Db 516 NNNNNNGGQTSMGMHPFYGNPNSAYGIILKBPDIHEYDEAKIDIGTFAQMIQATMGSS 575  
QY 144 NNATANQY 151  
Db 576 QGFNASAY 583

RESULT 12  
AD3143  
conserved hypothetical protein Atu4768 [imported] - Agrobacterium tumefaciens (strain C58)  
C:Species: Agrobacterium tumefaciens  
C>Date: 11-Jan-2002 #sequence\_revision 11-Jan-2002 #text\_change 18-Nov-2002  
C:Accession: AD3143  
R:Wood, D.W.; Setubal, J.C.; Kaul, R.; Monks, D.; Chen, L.; Wood, G.E.; Chen, Y.; Woo, L.;  
erage, G.; Gillet, W.; Grant, C.; Guenther, D.; Kutayavin, T.; Levy, R.; Li, M.; McClell  
; Karp, P.; Romero, P.; Zhang, S.  
Science 294, 2317-2323, 2001  
A:Authors: Yoo, H.; Tao, Y.; Biddle, P.; Jung, M.; Krespan, W.; Perry, M.; Gordon-Kamm,  
ster, E.W.  
A:Title: The Genome of the Natural Genetic Engineer Agrobacterium tumefaciens C58.  
A:Reference number: AB2577; MUID:21608550; PMID:11743193  
A:Accession: AD3143  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-145 <KUR>  
A:Cross-references: GB:AE008689; PIDN:AAL45562.1; PID:gl17743277; GSPDB:GN00187  
A:Experimental source: strain C58 (Dupont)  
C:Genetics:  
A:Gene: Atu4768  
A:Map position: linear chromosome

Query Match 12.3%; Score 96; DB 2; Length 145;  
Best Local Similarity 27.1%; Pred. No. 0.41; 59; Indels 20; Gaps 5;  
Matches 38; Conservative 23; Mismatches 59;

QY 3 LLKVAAPAAIVVSGSALAGVVP-----QWG-----GGGNHGGGN-----SSGPDST 44  
Db 1 MIRKSFASALVALVGLSAAAPAMANDVRIEQYGWSNSAGGAGYGNRIPTYQGGYNR 60

QY 45 LSIYQGSANALALQSDARKSETTITQSGYNGADVGCGADNSTIETQNGFRNATID 104  
Db 61 IVGHQYGRHNLG-AVGQEGHDNYGTTQNGNRNAGIGQFGSNHTTILTDGNGNIAAGV 119

QY 105 QWNAKNSDITVGYGGNNAA 124  
Db 120 Q-VGRGCSANVSQGGNDNVA 138

RESULT 13  
H98144  
hypothetical protein AGR\_L\_228 [imported] - Agrobacterium tumefaciens (strain C58), Cere  
C:Species: Agrobacterium tumefaciens  
C>Date: 22-Oct-2001 #sequence\_revision 22-Oct-2001 #text\_change 18-Nov-2002  
C:Accession: H98144  
R:Goodner, B.; Hinkle, G.; Gattung, S.; Miller, N.; Blanchard, M.; Qurollo, B.; Goldman,  
A.; Liu, F.; Wollam, C.; Allinger, M.; Doughty, D.; Scott, C.; Lappas, C.; Markelz, B.;  
Science 294, 2323-2328, 2001  
A:Title: Genome Sequence of the Plant Pathogen and Biotechnology Agent Agrobacterium tum  
A:Reference number: A97359; MUID:21608551; PMID:11743194  
A:Accession: H98144  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-145 <KUR>  
A:Cross-references: GB:AE007870; PIDN:AAK8682.1; PID:gl15158413; GSPDB:GN00170  
C:Genetics:  
A:Gene: AGR\_L\_228  
A:Map position: linear chromosome

Query Match 12.3%; Score 96; DB 2; Length 145;  
Best Local Similarity 27.1%; Pred. No. 0.41;  
Matches 38; Conservative 23; Mismatches 59; Indels 20; Gaps 5;

QY 3 LLKVAAPAAIVVSGSALAGVVP-----QWG-----GGGNHGGGN-----SSGPDST 44  
Db 1 MIRKSFASALVALVGLSAAAPAMANDVRIEQYGWSNSAGGAGYGNRIPTYQGGYNR 60

QY 45 LSIYQGSANALALQSDARKSETTITQSGYNGADVGCGADNSTIETQNGFRNATID 104  
Db 61 IVGHQYGRHNLG-AVGQEGHDNYGTTQNGNRNAGIGQFGSNHTTILTDGNGNIAAGV 119

QY 105 QWNAKNSDITVGYGGNNAA 124  
Db 120 Q-VGRGCSANVSQGGNDNVA 138

RESULT 14  
A70762  
probable PPE protein - Mycobacterium tuberculosis (strain H37RV)  
C:Species: Mycobacterium tuberculosis  
C>Date: 17-Jul-1998 #sequence\_revision 17-Jul-1998 #text\_change 15-Sep-2003  
C:Accession: A70762  
R:Cols, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon, S.;  
; Connor, R.; Davies, R.; Devlin, K.; Feltwell, T.; Gentles, S.; Hamlin, N.; Holroyd, S.;  
Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S.  
Nature 393, 537-544, 1998  
A:Authors: Sgares, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.  
A:Title: Deciphering the biology of Mycobacterium tuberculosis from the complete genome s  
A:Reference number: A70500; MUID:98295987; PMID:9634230  
A:Accession: A70762  
A:Status: preliminary; nucleic acid sequence not shown; translation not shown  
A:Molecule type: DNA  
A:Residues: 1-678 <COL>  
A:Cross-references: GB:Z74020; GB:AL123456; NID:g3261584; PIDN:CAA98335.1; PID:el300075;  
A:Experimental source: strain H37RV  
C:Genetics:  
A:Gene: PPE

Query Match 12.1%; Score 94.5; DB 2; Length 678;  
Best Local Similarity 25.7%; Pred. No. 3.1;  
Matches 39; Conservative 24; Mismatches 48; Indels 41; Gaps 8;

QY 15 SCSALAGVVPQWGGGNGHNGGNSGSDPSTLSIYQGSANALALQ--SDARKSETTITQ 72  
Db 384 SSGNLG-----FGNSGNIGFFNSG-NNNIGMNSGNGVGSALSVFSGSAERS----- 432

QY 73 SGYNGADVGCGADNS-----TIELTQNGFRNATIDQ--WNAKNSDITVGYGGNN 122  
Db 433 SGFGNSGELSTGIGNSGQLSTGWFNSATTTGWFNSGTTTGWFSGTTTGTGNSGSGN- 491

QY 123 AALVNYDQLVTRVVTHEM-----AHANNATAN 149  
Db 492 -----LVTGSMGLFNSGHTNTGSPN 511

RESULT 15  
AH0038  
probable exported protein YP00309 [imported] - Yersinia pestis (strain CO92)  
C:Species: Yersinia pestis  
C>Date: 02-Nov-2001 #sequence\_revision 02-Nov-2001 #text\_change 02-Nov-2001  
C:Accession: AH0038  
R:Parkhill, J.; Wren, B.W.; Thomson, N.R.; Titball, R.W.; Holden, M.T.G.; Prentice, M.B.;  
deno-Tarraga, A.M.; Chillingworth, T.; Cronin, A.; Davies, R.M.; Davis, P.; Dougan, G.; F  
il, M.; Rutherford, K.; Simmonds, M.; Skelton, J.; Stevens, K.; Whitehead, S.; Barrell, F  
Nature 413, 523-527, 2001  
A:Title: Genome sequence of Yersinia pestis, the causative agent of plague.  
A:Reference number: AB0001; MUID:21470413; PMID:11586360  
A:Accession: AH0038  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-1238 <KUR>  
A:Cross-references: GB:AL590842; PIDN:CAC89171.1; PID:gl5978410; GSPDB:GN00175  
C:Genetics:  
A:Gene: YP00309

Query Match 12.0%; Score 93.5; DB 2; Length 1238;  
Best Local Similarity 25.1%; Pred. No. 7.3;  
Matches 45; Conservative 19; Mismatches 48; Indels 67; Gaps 9;  
Qy 7 AAFPAIVVSGSALAGVVPQWGGG---NHNGGN-----SSG-----P 41  
Db 409 SAFSAITATGHLTA---EWQQAMLQTHSSGLDATTLIHFNDITAMSSGISLINEANQG 464  
Qy 42 DSTLSIYQGSANAA-----LALQSDARKSETTITQSGYGN 77  
Db 465 TSTADITVTCQINVSHEGITLNAITDGRTLNVVNNIASEYDAIRLYNYNDNYAT 524  
Qy 78 GADVQOGADN--STIEL-----TONGFENNATIDQWNAKNSDITVG--QYGGNNAAL 125  
Db 525 GVDDGTGADNGTSTIDLITREGALVSOQGYGINI---ETNTADTYVTVGGVLVHGGNGTAI 580

Search completed: August 2, 2004, 14:56:21  
Job time : 10.4 secs



GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: August 2, 2004, 14:36:12 ; Search time 5.3 Seconds  
(without alignments)

1483.508 Million cell updates/sec

Title: US-09-543-407-12

Perfect score: 779

Sequence: 1 MKLLKVAAPAAIVVSGSALA.....VTRVTHMAHANNATANQY 151

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 141681 seqs, 52070155 residues

Total number of hits satisfying chosen parameters: 141681

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt\_42.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	695	89.2	151	1	CSGA_SALTY
2	520	66.8	151	1	CSGA_ECOLI
3	501.5	64.4	152	1	CSGA_ECO57
4	101.5	13.0	151	1	CSGB_SALTY
5	100.5	12.9	151	1	CSGB_SALTY
6	98.5	12.6	1028	1	OVO_DROME
7	97	12.5	1656	1	OMP5_RICJA
8	94.5	12.1	1678	1	YF48_MYCTU
9	92	11.8	1327	1	TNKL_HUMAN
10	91.5	11.7	151	1	CSGB_ECOLI
11	89.5	11.5	262	1	VG38_BPT2
12	89	11.4	347	1	MSA2_PLAF2
13	88	11.3	491	1	YK98_MYCTU
14	87	11.2	331	1	OMB2_NEIMB
15	86	11.0	1093	1	PER_DROME
16	85	10.9	1250	1	YF48_ECOLI
17	84.5	10.8	3590	1	FFAB_BORPE
18	84	10.8	566	1	AMY_STRIM
19	84	10.8	566	1	AMY_STRIM
20	83.5	10.7	1067	1	SGG_DROME
21	83	10.7	590	1	GP63_LEIDO
22	83	10.7	599	1	GP63_LEICH
23	83	10.7	602	1	GP63_LEIMA
24	83	10.7	646	1	GP63_LEIMA
25	82.5	10.6	369	1	PST3_MYCAP
26	82.5	10.6	794	1	YC84_MYCPN
27	82	10.5	1567	1	ICEN_XANCT
28	81.5	10.5	392	1	HME1_HUMAN
29	81.5	10.5	730	1	GLN3_YEAST
30	81	10.4	165	1	PER_ORYSA
31	81	10.4	385	1	GRP1_DROME
32	81	10.4	485	1	YB47_MYCPN
33	81	10.4	959	1	N100_YEAST

#### ALIGNMENTS

##### RESULT 1

ID	CSGA_SALTY	STANDARD;	PRT;	151 AA.
AC	P55225;			
DT	01-OCT-1996 (Rel. 34, Created)			
DT	01-OCT-1996 (Rel. 34, Last sequence update)			
DT	10-OCT-2003 (Rel. 42, Last annotation update)			
DE	Major curlin subunit precursor (Fimbrin SEF17).			
GN	CSGA OR AGFA OR STM1144 OR STV1181 OR T1776.			
OS	Salmonella typhimurium,			
OS	Salmonella typhi, and			
OC	Salmonella enteritidis.			
OC	Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;			
OC	Enterobacteriaceae; Salmonella.			
OX	NCBI_TaxID=602, 601, 592;			
RN	[1]_TaxID=602, 601, 592;			
RP	SEQUENCE FROM N.A.			
RC	SPECIES=S.typhimurium; STRAIN=SR-11;			
RX	MEDLINE=98117058; PubMed=9457880;			
RA	Romling U., Bian Z., Hammar M., Sierralta W.D., Normark S.;			
RT	"Curli fibers are highly conserved between Salmonella typhimurium and			
RT	Escherichia coli with respect to operon structure and regulation.";			
RL	J. Bacteriol. 180:722-731(1998).			
RN	[2]			
RP	SEQUENCE FROM N.A.			
RC	SPECIES=S.typhimurium; STRAIN=LT2 / SGSC1412 / ATCC 700720;			
RX	MEDLINE=21534948; PubMed=11677609;			
RA	McClelland M., Sanderson K.E., Spieth J., Clifton S.W., Latreille P.,			
RA	Courtney L., Porwollik S., Ali J., Dante M., Du F., Hou S., Layman D.,			
RA	Leonard S., Nguyen C., Scott K., Holmes A., Grewal N., Mulvaney E.,			
RA	Ryan E., Sun H., Florea L., Miller W., Stoneking T., Nhan M.,			
RA	Waterston R., Wilson R.K.;			
RT	"Complete genome sequence of Salmonella enterica serovar Typhimurium			
RT	LT2.";			
RL	Nature 413:852-856(2001).			
RN	[3]			
RP	SEQUENCE FROM N.A.			
RC	SPECIES=S.typhi; STRAIN=CT18;			
RX	MEDLINE=21534947; PubMed=11677608;			
RA	Parkhill J., Dougan G., James K.D., Thomson N.R., Pickard D., Wain J.,			
RA	Churcher C., Mungall K.L., Bentley S.D., Holden M.T.G., Sebaihia M.,			
RA	Baker S., Basham D., Brooks K., Chillingworth T., Connor P.,			
RA	Cronin A., Davis P., Davies R.M., Dowd L., White N., Farrar J.,			
RA	Feltwell T., Hamlin N., Haque A., Hien T.T., Holroyd S., Jagels K.,			
RA	Krogan A., Larsen T.S., Leather S., Moule S., O'Gaora P., Parry C.,			
RA	Quail M.A., Rutherford K., Simmonds M., Skelton J., Stevens K.,			
RA	Whitehead S., Barrall B.G.;			
RT	"Complete genome sequence of a multiple drug resistant Salmonella			
RT	enterica serovar Typhi CT18.";			
RL	Nature 413:848-852(2001).			
RN	[4]			
RP	SEQUENCE FROM N.A.			
RC	SPECIES=S.typhi; STRAIN=Ty2 / ATCC 700931;			
RX	MEDLINE=22531367; PubMed=12644504;			
RA	Deng W., Liou S.-R., Plunkett G. III, Mayhew G.F., Rose D.J.,			
RA	Burland V., Kodyianni V., Schwartz D.C., Blattner F.R.;			

Q9pjy1 chlamydia m  
P42611 mycobacteri  
P33666 escherichia  
O50367 mycoplasma  
P28159 drosophila  
Q8in81 drosophila  
Q9utk4 schizosacch  
O83172 treponema p  
P75780 escherichia  
P27053 campylobact  
O8ixf0 homo sapien  
Q02630 saccharomyc

34 81 10.4 1460 1 PMPC\_CHLMU  
35 80.5 10.3 487 1 Y442\_MYCTU  
36 80.5 10.3 2003 1 YDBA\_ECOLI  
37 80 10.3 362 1 P35\_MYCPE  
38 80 10.3 594 1 SUH\_DROME  
39 80 10.3 955 1 FRU\_DROME  
40 80 10.3 1778 1 N189\_SCHPO  
41 79.5 10.2 485 1 Y136\_TREPA  
42 79.5 10.2 760 1 YBIL\_ECOLI  
43 79 10.1 572 1 FLAA\_CAMGO  
44 79 10.1 933 1 NPA3\_HUMAN  
45 79 10.1 1113 1 N116\_YEAST

RT "Comparative genomics of Salmonella enterica serovar Typhi strains Ty2  
and CT18.";  
J. Bacteriol. 185:2330-2337(2003).  
[5]  
SEQUENCE FROM N.A.  
RC SPECIES=S.enteritidis; STRAIN=27655-3B;  
RX MEDLINE=96146512; PubMed=8550497;  
RA Collinson S.K., Clouthier S.C., Doran J.L., Baner P.A., Kay W.W.;  
RT "Salmonella enteritidis agfBAC operon encoding thin, aggregative  
fimbriae.";  
J. Bacteriol. 178:662-667(1996).  
[6]  
SEQUENCE OF 21-151 FROM N.A.  
RC SPECIES=S.enteritidis; STRAIN=27655-3B;  
RX MEDLINE=94013373; PubMed=8104955;  
RA Doran J.L., Collinson S.K., Burian J., Santos G., Todd E.C.D.,  
Munro C.K., Kay C.M., Baner P.A., Peterkin P.I., Kay W.W.;  
RT "DNA-based diagnostic tests for Salmonella species targeting agfA,  
the structural gene for thin, aggregative fimbriae.";  
J. Clin. Microbiol. 31:2263-2273(1993).  
[7]  
SEQUENCE OF 21-33.  
RC SPECIES=S.enteritidis; STRAIN=27655-3B;  
RX MEDLINE=9130586; PubMed=1677357;  
RA Collinson S.K., Emeody L., Mueller K.-M., Trust T.J., Kay W.W.;  
RT "Purification and characterization of thin, aggregative fimbriae from  
Salmonella enteritidis.";  
J. Bacteriol. 173:4773-4781(1991).  
CC -1- FUNCTION: CURLIN IS THE STRUCTURAL SUBUNIT OF THE CURLI. CURLI ARE  
COILED SURFACE STRUCTURES THAT ASSEMBLE PREFERENTIALLY AT GROWTH  
TEMPERATURES BELOW 37 DEGREES CELSIUS. CURLI CAN BIND TO  
FIBRONECTIN.  
-1- SIMILARITY: BELONGS TO THE CSGA/CSGB FAMILY.  
-----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
between the Swiss Institute of Bioinformatics and the EMBL outstation -  
the European Bioinformatics Institute. There are no restrictions on its  
use by non-profit institutions as long as its content is in no way  
modified and this statement is not removed. Usage by and for commercial  
entities requires a license agreement (see <http://www.isb-sib.ch/announce/>  
or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
-----  
CC EMBL; AJ002301; CAA05317.1; -;  
DR EMBL; AB008749; AAL20074.1; -;  
DR EMBL; AL627269; CAD08268.1; -;  
DR EMBL; AE016840; AA069399.1; -;  
DR EMBL; U43280; AAC43599.1; -;  
DR PIR; JC6039; JC6039.  
DR StyGene; SG10608; csga.  
KW Fimbria; Signal; Complete proteome.  
FT SIGNAL 1 20  
FT CHAIN 21 151 MAJOR CURLIN SUBUNIT.  
FT CONFLICT 134 151 SVMYRQVGFNGNATANY -> DSYTQVAS (IN  
REF. 6).  
FT SEQUENCE 151 AA; 15305 MW; B7DAC0D16B621359 CRC64;  
-----  
Query Match 89.2%; Score 695; DB 1; Length 151;  
Best Local Similarity 90.1%; Pred. No. 4.6e-51;  
Matches 136; Conservative 4; Mismatches 11; Indels 0; Gaps 0;  
-----  
QY 1 MKLLKVAFAAIVVSGSALAGVFWQGGGNNHNGGSSGPDSTLSIYQGSANAALALQ 60  
DB 1 MKLLKVAFAAIVVSGSALAGVFWQGGGNNHNGGSSGPDSTLSIYQGSANAALALQ 60  
-----  
QY 61 SDARKSETTITQSGVNGADVGQADNSTIELTQNGFRFNATIDQWNAKNSDITVGYGG 120  
DB 61 SDARKSETTITQSGVNGADVGQADNSTIELTQNGFRFNATIDQWNAKNSDITVGYGG 120  
-----  
QY 121 NNAALVNDQLVTRVVTTHMAHANNATANY 151  
DB 121 NNAALVNDQLVTRVVTTHMAHANNATANY 151

RESULT 2  
CSGA\_ECOLI STANDARD; PRT; 151 AA.  
AC P28307;  
DT 01-DEC-1992 (Rel. 24, Created)  
DT 01-OCT-1996 (Rel. 34, Last sequence update)  
DT 28-FEB-2003 (Rel. 41, Last annotation update)  
DE Major curlin subunit precursor.  
GN CSGA OR B1042.  
OS Escherichia coli.  
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;  
OC Enterobacteriaceae; Escherichia.  
OX NCBI\_TaxID=562;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=K12 / W3110;  
RX MEDLINE=93211294; PubMed=8459772;  
RA Olsen A., Arngvist A.;  
RT "The rpoS sigma factor relieves H-NS-mediated transcriptional  
repression of csga, the subunit gene of fibronectin-binding curli in  
Escherichia coli.";  
Mol. Microbiol. 7:523-536(1993).  
RN [2]  
RP SEQUENCE FROM N.A.  
RC STRAIN=K12 / MC4100;  
RX MEDLINE=96414468; PubMed=8817489;  
RA Hammar M., Arngvist A., Bian Z., Olsen A., Normark S.;  
RT "Expression of two csq operons is required for production of  
fibronectin- and congo red-binding curli polymers in Escherichia coli  
K-12.";  
Mol. Microbiol. 18:661-670(1995).  
RN [3]  
RP SEQUENCE FROM N.A.  
RC STRAIN=K12 / MG1655;  
RX MEDLINE=97426617; PubMed=9278503;  
RA Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,  
Riley M., Collado-Valdes J., Glasner J.D., Rode C.K., Mayhew G.F.,  
Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,  
Mau B., Shao Y.;  
RT "The complete genome sequence of Escherichia coli K-12.";  
Science 277:1453-1474(1997).  
RN [4]  
RP SEQUENCE FROM N.A.  
RC STRAIN=K12;  
RX MEDLINE=97061202; PubMed=8905232;  
RA Oshima T., Aiba H., Baba T., Fujita K., Hayashi K., Honjo A.,  
Ikemoto K., Inada T., Itoh T., Kajihara M., Kanai K., Kashimoto K.,  
Kimura S., Kitagawa M., Makino K., Masuda S., Miki T., Mizobuchi K.,  
Mori H., Motomura K., Nakamura Y., Washimoto H., Nishio Y., Saito N.,  
Sampei G., Seki Y., Tagami H., Takemoto K., Wada C., Yamamoto Y.,  
Yano M., Horiuchi T.;  
RT "A 718-kb DNA sequence of the Escherichia coli K-12 genome  
corresponding to the 12.7-28.0 min region on the linkage map.";  
DNA Res. 3:137-155(1996).  
RN [5]  
RP SEQUENCE OF 21-40.  
RC STRAIN=K12 / YWEL;  
RX MEDLINE=93023873; PubMed=1357528;  
RA Arngvist A., Olsen A., Pfeiffer J., Russell D.G., Normark S.;  
RT "The Crl protein activates cryptic genes for curli formation and  
fibronectin binding in Escherichia coli HB101.";  
Mol. Microbiol. 6:2443-2452(1992).  
RN [6]  
RP SEQUENCE OF 21-31.  
RX MEDLINE=9130586; PubMed=1677357;  
RA Collinson S.K., Emeody L., Trust T.J., Kay W.W.;  
RT "Purification and characterization of thin, aggregative fimbriae from  
Salmonella enteritidis.";  
J. Bacteriol. 173:4773-4781(1991).  
CC -1- FUNCTION: CURLIN IS THE STRUCTURAL SUBUNIT OF THE CURLI. CURLI ARE  
COILED SURFACE STRUCTURES THAT ASSEMBLE PREFERENTIALLY AT GROWTH  
TEMPERATURES BELOW 37 DEGREES CELSIUS. CURLI CAN BIND TO  
FIBRONECTIN.

```

CC      -!- SIMILARITY: BELONGS TO THE CSGA/CSGB FAMILY.
CC      This SWISS-PROT entry is copyright. It is produced through a collaboration
CC      between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC      the European Bioinformatics Institute. There are no restrictions on its
CC      use by non-profit institutions as long as its content is in no way
CC      modified and this statement is not removed. Usage by and for commercial
CC      entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC      or send an email to license@isb-sib.ch).
CC      -----
CC      EMBL; L04979; AAA23616.1; --
CC      EMBL; X90754; CAA62282.1; --
CC      EMBL; AE000205; AAC74126.1; --
CC      EMBL; D90741; BAA35832.1; --
CC      EMBL; D90742; BAA35840.1; --
CC      PIR; S70788; S70788.
CC      EcoGene; EG11489; csga.
CC      FimBria; Signal; Complete proteome.
CC      FT SIGNAL 1 20 MAJOR CURLIN SUBUNIT.
CC      FT CHAIN 21 151
CC      FT CONFLICT 27
CC      FT SEQUENCE 151 AA; 15049 MW; C003470D208D395F CRC64;
CC      -----
CC      Query Match 66.8%; Score 520; DB 1; Length 151;
CC      Best Local Similarity 66.9%; Pred. No. 1.5e-36;
CC      Matches 101; Conservative 21; Mismatches 29; Indels 0; Gaps 0;
CC      -----
CC      QY 1 MKLLKVAFAAIVVSGSALAGVFPQWGGGNGHNGGNSGPDSTLSIYQYGSANAALALQ 60
CC      Db EMBL; AP002554; BAB34843.1; --
CC      1 MKLLKVAFAAIVVSGSALAGVFPQYGGGNGHNGGNSGPNSELNIYQYGGNSALALQ 60
CC      -----
CC      QY 61 SDARKSETTITQSGYNGADVGGADNNTIELTQNGFRNATIDQWNAKNSDITVGYG 120
CC      Db TDRNSDLTITQGGNGADVGGSDSSDLTQRFSGSATLDQWNGKNSMTVKQFG 120
CC      -----
CC      QY 121 NNAALVNYDOLVTRVTHEVAHANNATANOY 151
CC      Db 121 GNGAAVDQTASNSVNTVQVFGNNATAHOY 151
CC      -----
CC      RESULT 3
CC      CSGA_ECO57 STANDARD; PRT; 152 AA.
CC      ID CSGA_ECO57 STANDARD; PRT; 152 AA.
CC      AC Q93U24;
CC      DT 28-FEB-2003 (Rel. 41, Created)
CC      DT 28-FEB-2003 (Rel. 41, Last sequence update)
CC      DT 28-FEB-2003 (Rel. 41, Last annotation update)
CC      DE Major curlin subunit precursor.
CC      GN Escherichia coli O157:H7.
CC      OS Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
CC      OC Enterobacteriaceae; Escherichia.
CC      OX NCBI_TaxID=83334;
CC      RN [1]
CC      RP SEQUENCE FROM N.A.
CC      RC STRAIN=0157:H7 / ATCC 43995;
CC      RX MEDLINE=21218556; PubMed=11319125;
CC      RA "Ulrich G.A., Keen J.E., Elder R.O.;
CC      RA Posfai G., Hackett J., Klink S., Boutin A., Shao Y., Miller L.,
CC      RA Grobeck E.J., Davis N.W., Lim A., Dimalanta E.T., Potamousis K.,
CC      RA Apodaca J., Anantharaman T.S., Lin J., Yen G., Schwartz D.C.,
CC      RA Welch R.A., Blatterer P.R.;
CC      RA "Genome sequence of enterohemorrhagic Escherichia coli O157:H7."
CC      RL Appl. Environ. Microbiol. 67:2367-2370(2001).
CC      RN [2]
CC      RP SEQUENCE FROM N.A.
CC      RC STRAIN=0157:H7 / BDL933 / ATCC 700927;
CC      RX MEDLINE=21074935; PubMed=11206551;
CC      RA Perna N.T., Plunkett G. III, Burland V., Mau B., Glasner J.D.,
CC      RA Rose D.J., Mayhew G.F., Evans P.S., Gregor J., Kirkpatrick H.A.,
CC      RA Posfai G., Hackett J., Klink S., Boutin A., Shao Y., Miller L.,
CC      RA Grobeck E.J., Davis N.W., Lim A., Dimalanta E.T., Potamousis K.,
CC      RA Apodaca J., Anantharaman T.S., Lin J., Yen G., Schwartz D.C.,
CC      RA Welch R.A., Blatterer P.R.;
CC      RA "Genome sequence of enterohemorrhagic Escherichia coli O157:H7."
CC      RL Nature 409:529-533(2001).
RN      [3]
RN      SEQUENCE FROM N.A.
RN      RC STRAIN=0157:H7 / RIMD 0509952;
RN      RX MEDLINE=21156231; PubMed=11258796;
RN      RA Hayashi T., Makino K., Ohnishi M., Kurokawa K., Ishii K., Yokoyama K.,
RN      RA Han C.-G., Onosubo E., Nakayama K., Murata T., Tanaka M., Tobe T.,
RN      RA Iida T., Takami H., Honda T., Sasaki C., Ogasawara N., Yasunaga T.,
RN      RA Kuhara S., Shiba T., Hattori M., Shinagawa H.;
RN      RT "Complete genome sequence of enterohemorrhagic Escherichia coli
RN      RT O157:H7 and genomic comparison with a laboratory strain K-12."
RN      RL DNA Res. 8:11-22(2001).
RN      CC -!- FUNCTION: CURLIN IS THE STRUCTURAL SUBUNIT OF THE CURLI. CURLI ARE
RN      CC COILED SURFACE STRUCTURES THAT ASSEMBLE PREFERENTIALLY AT GROWTH
RN      CC TEMPERATURES BELOW 37 DEGREES CELSIUS. CURLI CAN BIND TO
RN      CC FIBRONECTIN.
RN      CC -!- SIMILARITY: BELONGS TO THE CSGA/CSGB FAMILY.
RN      CC -----
RN      CC This SWISS-PROT entry is copyright. It is produced through a collaboration
RN      CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
RN      CC the European Bioinformatics Institute. There are no restrictions on its
RN      CC use by non-profit institutions as long as its content is in no way
RN      CC modified and this statement is not removed. Usage by and for commercial
RN      CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
RN      CC or send an email to license@isb-sib.ch).
RN      CC -----
RN      CC EMBL; AF275733; AAK53212.1; --
RN      CC EMBL; AE005315; AAG55788.1; --
RN      CC EMBL; AP002554; BAB34843.1; --
RN      CC PIR; D90806; D90806.
RN      CC PIR; H85665; H85665.
RN      KW FimBria; Signal; Complete proteome.
RN      FT SIGNAL 1 20 BY SIMILARITY.
RN      FT CHAIN 21 152 MAJOR CURLIN SUBUNIT.
RN      FT SEQUENCE 152 AA; 15099 MW; SE2D2D94DDE91243 CRC64;
RN      -----
RN      Query Match 64.4%; Score 501.5; DB 1; Length 152;
RN      Best Local Similarity 65.8%; Pred. No. 5.1e-35;
RN      Matches 100; Conservative 21; Mismatches 30; Indels 1; Gaps 1;
RN      -----
RN      QY 1 MKLLKVAFAAIVVSGSALAGVFPQWGGGNGHNGGNSGPDSTLSIYQYGSANAALAL 59
RN      Db 1 MKLLKVAFAAIVVSGSALAGVFPQYGGGNGHNGGNSGPNSELNIYQYGGNSALAL 60
RN      -----
RN      QY 60 QSDARKSETTITQSGYNGADVGGADNNTIELTQNGFRNATIDQWNAKNSDITVGYG 119
RN      Db 61 QADARNSDLTITQGGNGADVGGSDSSDLTQRFSGSATLDQWNGKNSMTVKQFG 120
RN      -----
RN      QY 120 GNGAAVDQTASNSVNTVQVFGNNATAHOY 151
RN      Db 121 GNGAAVDQTASNSVNTVQVFGNNATAHOY 152
RN      -----
RN      RESULT 4
RN      CSGB_SALTI STANDARD; PRT; 151 AA.
RN      ID CSGB_SALTI STANDARD; PRT; 151 AA.
RN      AC Q827M3;
RN      DT 28-FEB-2003 (Rel. 41, Created)
RN      DT 28-FEB-2003 (Rel. 41, Last sequence update)
RN      DT 10-OCT-2003 (Rel. 42, Last annotation update)
RN      DE Minor curlin subunit precursor.
RN      GN CSGB OR STY1180 OR T1777.
RN      OS Salmonella typhi.
RN      OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
RN      OC Enterobacteriaceae; Salmonella.
RN      OX NCBI_TaxID=601;
RN      RN [1]
RN      RP SEQUENCE FROM N.A.
RN      RC STRAIN=CT18;
RN      RX MEDLINE=21534947; PubMed=11677608;
RN      RA Parkhill J., Dougan G., James K.D., Thomson N.R., Pickard D., Wain J.,
RN      RA Churcher C., Mungall K.L., Bentley S.D., Holden M.T.G., Sebaihia M.,
RN      RA Baker S., Basham D., Brooks K., Chillingworth T., Connor P.,
RN      RA Cronin A., Davis P., Davies R.M., Dowd L., White N., Farrar J.,

```



RP SEQUENCE FROM N.A.  
RC TISSUE=Ovary;  
RX MEDLINE=95021209; PubMed=7935398;  
RA Garfinkel M.D., Wang J., Liang Y., Mahowald A.P.;  
RT "Multiple products from the shavenbaby-ovo gene region of Drosophila  
melanogaster: relationship to genetic complexity";  
RL Mol. Cell. Biol. 14:6809-6818(1994).  
RN [2].  
RP SEQUENCE FROM N.A.  
RC STRAIN=Oregon-R;  
RX MEDLINE=91293102; PubMed=1712294;  
RA Mevel-Ninio M.T.M., Terracol R., Kafatos F.C.;  
RT "The ovo gene of Drosophila encodes a zinc finger protein required  
for female germ line development";  
RL EMBO J. 10:2259-2266(1991).  
CC -!- FUNCTION: REQUIRED FOR SURVIVAL AND DIFFERENTIATION OF FEMALE GERM  
LINE CELLS. PLAYS A ROLE IN GERM LINE SEX DETERMINATION.  
CC -!- SUBCELLULAR LOCATION: Nuclear (Potential).  
CC -!- DEVELOPMENTAL STAGE: FIRST APPEARS IN THE GERMARIIUM AND  
ACCUMULATES IN NURSE CELLS DURING OOGENESIS. STORED IN THE EGG,  
BUT IS RAPIDLY LOST IN THE EMBRYOS EXCEPT FOR ITS CONTINUED  
PRESENCE IN THE GERM LINE PRECURSOR POLE CELLS.  
CC -!- SIMILARITY: Contains 4 C2H2-type zinc fingers.  
CC  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
between the Swiss Institute of Bioinformatics and the EMBL outstation -  
the European Bioinformatics Institute. There are no restrictions on its  
use by non-profit institutions as long as its content is in no way  
modified and this statement is not removed. Usage by and for commercial  
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC  
CC EMBL: U11383; AAB60216.1; -  
CC EMBL: X59772; CAB36921.1; ALT\_SEQ.  
CC PIR: A56038; A56038.  
CC HSP: P07248; ZADR.  
CC TRANSFAC: T00669; -  
CC FlyBase: FBgn003028; ovo.  
CC InterPro: IPR007087; Znf\_C2H2.  
CC Pfam: PF00096; zf-C2H2; 3.  
CC SMART: SM00355; Znf\_C2H2; 4.  
CC PROSITE: PS00028; ZINC\_FINGER\_C2H2\_1; 3.  
CC PROSITE: PS00157; ZINC\_FINGER\_C2H2\_2; 3.  
CC Zinc-finger; Metal-binding; DNA-binding; Repeat; Nuclear protein;  
KW Transcription regulation.  
FT DOMAIN 62 66 POLY-ALA.  
FT DOMAIN 72 77 POLY-GLY.  
FT DOMAIN 80 85 POLY-GLY.  
FT DOMAIN 98 108 POLY-GLY.  
FT DOMAIN 144 152 POLY-HIS.  
FT DOMAIN 153 159 POLY-ASN.  
FT DOMAIN 336 339 POLY-GLN.  
FT DOMAIN 347 353 POLY-GLN.  
FT DOMAIN 357 361 POLY-GLN.  
FT DOMAIN 410 414 POLY-GLN.  
FT DOMAIN 418 422 POLY-GLN.  
FT DOMAIN 426 432 POLY-GLN.  
FT DOMAIN 445 453 POLY-GLN.  
FT DOMAIN 456 459 POLY-GLN.  
FT DOMAIN 466 474 POLY-GLN.  
FT DOMAIN 497 517 POLY-ALA.  
FT DOMAIN 524 529 POLY-SER.  
FT DOMAIN 549 558 POLY-ALA.  
FT DOMAIN 639 651 POLY-ALA.  
FT DOMAIN 717 725 POLY-ALA.  
FT DOMAIN 797 802 POLY-GLN.  
FT DOMAIN 820 823 POLY-GLN.  
FT DOMAIN 826 832 POLY-GLN.  
FT ZN\_FING 874 896 C2H2-TYPE 1.  
FT ZN\_FING 902 924 C2H2-TYPE 2.  
FT ZN\_FING 930 953 C2H2-TYPE 3.  
FT ZN\_FING 969 992 C2H2-TYPE 4.  
FT CONFLICT 647 647 A -> R (IN REF. 2).

SQ SEQUENCE 1028 AA; 110620 MW; D7068BEB2BC0F6F77 CRC64;  
Query Match 12.6%; Score 98.5; DB 1; Length 1028;  
Best Local Similarity 24.5%; Pred. No. 1.1;  
Matches 46; Conservative 15; Mismatches 62; Indels 65; Gaps 8;  
QY 3 LLKVAFAAIVVSGSALAGVVPQWGGGNGHNGGNSGPDSTLSIYQVGSNAALALQSD 62  
DB 59 LQNAAAAYIMSAGS-----GGCTGNGGGAGSGPGGPPSANSGGGGGG----- 104  
QY 63 ARKSETTITQSGYNGADVGCGADNSTIELTQNGFRNNATIDQWNAKNSDI----- 113  
DB 105 -----GGGYINCAGVG-GPNN-----LDGNLLNFASVSNYNESKFNHHHHHHQ 152  
QY 114 -----TVGQ-----YGGNNAAL-----VNYDQ-----LVTRVVTHEMAHA 143  
DB 153 NNNNNNGGOTSMGHPFYGGNPSAYGIILKDEPDIEYDEAKIDIGTFAQNIIOATMGSS 212  
QY 144 NNATANQY 151  
DB 213 GQFNASAY 220  
RESULT 7  
ID OMPB RICJA STANDARD; PRT; 1656 AA.  
AC O06653;  
DT 30-MAY-2000 (Rel. 39, Created)  
DT 30-MAY-2000 (Rel. 39, Last sequence update)  
DT 16-OCT-2001 (Rel. 40, Last annotation update)  
DE Outer membrane protein B precursor (168 kDa surface-layer protein)  
DE (Surface protein antigen) (Cell surface antigen 5) (Scas) (rOmpB)  
DE (rOmp B) [Contains: 120 kDa surface-exposed protein (Surface protein  
antigen) (120 kDa outer membrane protein ompB); 32 kDa beta peptide].  
GN OMPB.  
OS Rickettsia japonica.  
OC Bacteria; Proteobacteria; Alphaproteobacteria; Rickettsiales;  
OC Rickettsiaceae; Rickettsiae; Rickettsia.  
OX NCBI\_TaxID=35790;  
RN [1]\_  
RP SEQUENCE FROM N.A.  
RC STRAIN=YH;  
RA Uchiyama T.;  
RT "Sequencing of the gene encoding the protein rOmp B of Rickettsia  
japonica";  
RL Submitted (MAY-1997) to the EMBL/GenBank/DBJ databases.  
CC -!- FUNCTION: THE 120 kDa SURFACE-EXPOSED PROTEIN IS A MAJOR  
STRUCTURAL PROTEIN WHICH MAY PLAY A ROLE AS A RICKETTSIAL  
VIRULENCE FACTOR AND/OR IMMUNOGEN DURING INFECTION (BY  
similarity).  
CC -!- FUNCTION: THE 32 kDa BETA PEPTIDE MAY SERVE AS A MEMBRANE ANCHOR  
(By similarity).  
CC -!- SUBCELLULAR LOCATION: Cell wall. This bacterium is covered by a s-  
layer with hexagonal symmetry.  
CC -!- SIMILARITY: BELONGS TO THE RICKETTSIAE OMPA/OMPB FAMILY.  
CC  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
between the Swiss Institute of Bioinformatics and the EMBL outstation -  
the European Bioinformatics Institute. There are no restrictions on its  
use by non-profit institutions as long as its content is in no way  
modified and this statement is not removed. Usage by and for commercial  
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC  
CC EMBL: AB003681; BAA20138.1; -  
CC InterPro: IPR006315; Autotransport.  
CC InterPro: IPR005546; Autotransporter.  
CC Pfam: PF03797; Autotransporter; 1.  
CC TIGRPFAMs: TIGR01414; autotrans\_bar1; 2.  
KW Antigen; S-layer; Cell wall.  
FT CHAIN 1 1338 120 kDa SURFACE-EXPOSED PROTEIN.  
FT CHAIN 1339 1656 32 kDa BETA PEPTIDE.  
FT DOMAIN 528 533 POLY-GLY.

SQ SEQUENCE 1656 AA; 168097 MW; 3132A69C9DD5999F CRC64;  
 Query Match 12.5%; Score 97; DB 1; Length 1656;  
 Best Local Similarity 28.3%; Pred. No. 2.6;  
 Matches 43; Conservative 19; Mismatches 52; Indels 38; Gaps 7;  
 QY 6 VAAFAIVVSGALAGVVPQWGGGNGGNSGPDSTLSIYQVGSANALALQSDARK 65  
 DB 509 VLAAGAILDGSATI-----TGDIAGGG-----GAALQSITLANDATK 547  
 QY 66 SETTITQSG-----YGNAGDVGGADNSTIETQNGFRNATIDQWNAKNSDITVGVG--QYG 119  
 DB 548 ---TLTGGANIISANGGTINFQANGGTTIKLTST--QNNIIVD-----CDLAIAITDQTG 596  
 QY 120 GNNALVNYDOLVTRVTHMAHANNATANCY 151  
 DB 597 VVDASLTNAQTLLTISGTIGANNITLQGF 628  
 RESULT 8  
 ID YF48 MYCTU STANDARD; PRT; 678 AA.  
 AC Q10778;  
 DT 01-OCT-1996 (Rel. 34, Created)  
 DT 15-JUL-1999 (Rel. 38, Last sequence update)  
 DT 10-OCT-2003 (Rel. 42, Last annotation update)  
 DE Hypothetical PPE-family protein RV1548C/MT1599.  
 GN RV1548C OR MT1599 OR MTCY48.17.  
 OS Mycobacterium tuberculosis.  
 OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;  
 OC Corynebacterineae; Mycobacteriaceae; Mycobacterium.  
 OX NCBI\_TaxID=1773;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=H37RV;  
 RX MEDLINE=98295987; PubMed=9634230;  
 RA Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,  
 RA Gordon S.V., Bigmeier K., Gas S., Barry C.E. III, Tekala F.,  
 RA Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,  
 RA Davies R., Devlin K., Feltwell T., Gentles S., Hamlin N., Holroyd S.,  
 RA Hornsby T., Jagels K., Krogh A., McLean J., Moule S., Murphy L.,  
 RA Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,  
 RA Rutter S., Seeger K., Skelton S., Squares S., Squares R.,  
 RA Sultston J.E., Taylor K., Whitehead S., Barrell B.G.;  
 RT "Deciphering the biology of Mycobacterium tuberculosis from the  
 RT complete genome sequence."  
 RL Nature 393:537-544(1998).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=CDC 1551 / Oshkosh;  
 RX MEDLINE=22206494; PubMed=12218036;  
 RA Fleischmann R.D., Alland D., Eisen J.A., Carpenter L., White O.,  
 RA Peterson J., DeBoy R., Dodson R., Gwinn M., Haft D., Hickey E.,  
 RA Kolonay J.F., Nelson W.C., Unwayan L.A., Ermolaeva M., Salzberg S.L.,  
 RA Delcher A., Utterback T., Weidman J., Khouri H., Gill J., Mikula A.,  
 RA Bishai W., Jacobs W.R. Jr., Venter J.C., Fraser C.M.;  
 RT "Whole-genome comparison of Mycobacterium tuberculosis clinical and  
 RT laboratory strains."  
 RL J. Bacteriol. 184:5479-5490(2002).  
 RN [3]  
 RP SIMILARITY: Belongs to the mycobacterial PPE family.  
 CC  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC  
 CC EMBL; Z74020; CAA98335.1;  
 DR EMBL; A007026; AAK45866.1; ALT\_INIT.  
 DR PIR; A70762; A70762.  
 DR TIGR; MT1599; --

Tuberculin; RV1548C;  
 DR InterPro; IPR000030; Microbac\_PPE.  
 DR InterPro; IPR002989; Mycobac\_Pentapep.  
 DR Pfam; PF01469; Pentapeptide\_2; 11.  
 DR Pfam; PF00823; PPE; 1.  
 KW Hypothetical protein; Transmembrane; Complete proteome.  
 FT TRANSMEM 14 34 POTENTIAL.  
 FT TRANSMEM 180 200 POTENTIAL.  
 FT CONFLICT 258 258 D -> G (IN REF. 2).  
 SQ SEQUENCE 678 AA; 66736 MW; 209F1593D52533A2 CRC64;  
 Query Match 12.1%; Score 94.5; DB 1; Length 678;  
 Best Local Similarity 25.7%; Pred. No. 1.5;  
 Matches 39; Conservative 24; Mismatches 48; Indels 41; Gaps 8;  
 QY 15 SGSALAGVVPQWGGGNGGNSGPDSTLSIYQVGSANALALQ--SDARKSETTITQ 72  
 DB 384 SGSGNLG-----FGNSGNGNIGFFNSG--NNNIGMNSGNGVGALSVEFGSSAERS----- 432  
 QY 73 SGYNGADVGGADNS-----TIELTQNGFRNATIDQ--WNAKNSDITVGVGGNN 122  
 DB 433 SGFGNSGELSTGICNSGQLSTGWFNSATSTGWFNSGTTNTGKNSGTTNTGKNSGNN- 491  
 QY 123 AALVNDQLVTRVVTHEM-----AHANNATAN 149  
 DB 492 -----LVTGSMGLFNSGHTNTGTFN 511  
 RESULT 9  
 ID TNK1 HUMAN STANDARD; PRT; 1327 AA.  
 AC Q95271; Q95272;  
 DT 28-FEB-2003 (Rel. 41, Created)  
 DT 28-FEB-2003 (Rel. 41, Last sequence update)  
 DT 10-OCT-2003 (Rel. 42, Last annotation update)  
 DE Tankyrase 1 (EC 2.4.2.30) (TANKI) (Tankyrase I) (TNKS-1) (TRF1-  
 DE interacting ankyrin-related ADP-ribose polymerase).  
 GN TNKS OR TNKSI OR TIN1 OR TINP1 OR PARP1.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Eukarya; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Theria; Primates; Catarrhini; Hominidae; Homo.  
 OX NCBI\_TaxID=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A. (ISOFORMS 1 AND 2).  
 RC TISSUE=Testis;  
 RX MEDLINE=99040105; PubMed=9822378;  
 RA Smith S., Giriat L., Schmitt A., de Lange T.;  
 RT "Tankyrase, a poly(ADP-ribose) polymerase at human telomeres."  
 RL Science 282:1484-1487(1998).  
 RN [2]  
 RP SUBCELLULAR LOCATION.  
 RX MEDLINE=99454782; PubMed=10523501;  
 RA Smith S., de Lange T.;  
 RT "Cell cycle dependent localization of the telomeric PARP, tankyrase,  
 RT to nuclear pore complexes and centrosomes."  
 RL J. Cell Sci. 112:3649-3656(1999).  
 RN [3]  
 RP FUNCTION, AND PHOSPHORYLATION.  
 RX MEDLINE=20536282; PubMed=10988299;  
 RA Chi N.-W., Lodish H.F.;  
 RT "Tankyrase is a Golgi-associated mitogen-activated protein kinase  
 RT substrate that interacts with TRAP in GLUT4 vesicles."  
 RL J. Biol. Chem. 275:38437-38444(2000).  
 RN [4]  
 RP FUNCTION, AND MUTAGENESIS OF HIS-1184 AND GLU-1291.  
 RX MEDLINE=21602874; PubMed=11739745;  
 RA Cook B.D., Dymek J.N., Chang W., Shostak G., Smith S.;  
 RT "Role for the related poly(ADP-Ribose) polymerases tankyrase 1 and 2  
 RT at human telomeres."  
 RL Mol. Cell. Biol. 22:332-342(2002).  
 CC -!- FUNCTION: May regulate vesicle trafficking and modulate the  
 CC subcellular distribution of SLC24A/GLUT4-vesicles. Has PARP  
 CC activity and can modify TRF1, and thereby contribute to the

regulation of telomere length.  
 CC -!- CATALYTIC ACTIVITY: NAD(+) + {ADP-D-riboseyl}{N}-acceptor =  
 CC nicotinamide + {ADP-D-riboseyl}{N+1}-acceptor.  
 CC -!- SUBUNIT: Oligomerizes and associates with TNKS2. Interacts with  
 CC the cytoplasmic domain of INPEP/Otase in SLC2A4/GLUT4-vesicles.  
 CC Binds to the N-terminus of telomeric TRF1 via the ANK repeats.  
 CC -!- SUBCELLULAR LOCATION: Cytoplasmic, associated with the Golgi and  
 CC with juxtanuclear SLC2A4/GLUT4-vesicles. A minor proportion is  
 CC also found at nuclear pore complexes and around the pericentriolar  
 CC matrix of mitotic centrosomes. During interphase, a small fraction  
 CC of TNKS is found in the nucleus, associated with TRF1.  
 CC -!- ALTERNATIVE PRODUCTS:  
 CC Event-Alternative splicing; Named isoforms=2;  
 CC Name=1;  
 CC IsoId=O95271-1; Sequence=displayed;  
 CC Name=2;  
 CC IsoId=O95271-2; Sequence=VSP\_004538, VSP\_004539;  
 CC Note=No experimental confirmation available;  
 CC -!- TISSUE SPECIFICITY: Ubiquitous; highest levels in testis.  
 CC -!- PTM: Upon insulin-stimulation, phosphorylated on serine residues  
 CC by MAPK kinases.  
 CC -!- PTM: ADP-ribosylated (-auto).  
 CC -!- SIMILARITY: Belongs to the PARP family.  
 CC -!- SIMILARITY: Contains 15 ANK repeats.  
 CC -!- SIMILARITY: Contains 1 sterile alpha motif (SAM) domain.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>)  
 CC or send an email to [license@sib-sib.ch](mailto:license@sib-sib.ch).  
 CC -----  
 CC DR EMBL; AF082556; AAC79841.1; -  
 CC DR EMBL; AF082557; AAC79842.1; -  
 CC DR EMBL; AF082558; AAC79843.1; -  
 CC DR EMBL; AF082559; AAC79844.1; -  
 CC DR HSSP; Q00420; IAWC.  
 CC DR Genew; HGNC:11941; TNKS.  
 CC DR MIM; 603303; -  
 CC DR GO; GO:0000781; C:chromosome, telomeric region; IDA.  
 CC DR GO; GO:0003950; F:NAD ADP-ribosyltransferase activity; IDA.  
 CC DR GO; GO:0005515; F:protein binding; IPI.  
 CC DR GO; GO:0007004; P:telomerase-dependent telomere maintenance; NAS.  
 CC DR InterPro; IPR002110; ANK.  
 CC DR InterPro; IPR001660; SAM.  
 CC DR Pfam; PF00023; ank; 19.  
 CC DR Pfam; PF00536; SAM; 1.  
 CC DR PRINTS; PR01415; ANKYRIN.  
 CC DR SMART; SM00248; ANK; 17.  
 CC DR SMART; SM00454; SAM; 1.  
 CC DR PROSITE; PS50088; ANK\_REPEAT; 15.  
 CC DR PROSITE; PS50297; ANK\_REPEAT\_REGION; 1.  
 CC DR PROSITE; PS50105; SAM\_DOMAIN; 1.  
 CC KW Transferrase; Glycosyltransferase; NAD; Golgi stack; Telomere;  
 CC Nuclear protein; Repeat; ANK repeat; ADP-ribosylation;  
 CC Phosphorylation; Alternative splicing.  
 CC FT REPEAT 215 247 ANK 1.  
 CC FT REPEAT 248 280 ANK 2.  
 CC FT REPEAT 281 313 ANK 3.  
 CC FT REPEAT 368 400 ANK 4.  
 CC FT REPEAT 401 433 ANK 5.  
 CC FT REPEAT 434 466 ANK 6.  
 CC FT REPEAT 521 556 ANK 7.  
 CC FT REPEAT 557 589 ANK 8.  
 CC FT REPEAT 590 622 ANK 9.  
 CC FT REPEAT 683 715 ANK 10.  
 CC FT REPEAT 716 748 ANK 11.  
 CC FT REPEAT 749 781 ANK 12.  
 CC FT REPEAT 836 868 ANK 13.  
 CC FT REPEAT 869 901 ANK 14.  
 CC FT REPEAT 902 934 ANK 15.

FT DOMAIN 1030 1089 SAM.  
 FT DOMAIN 1176 1327 PARP.  
 FT DOMAIN 9 14 POLY-HIS.  
 FT DOMAIN 27 34 POLY-PRO.  
 FT DOMAIN 128 134 POLY-SER.  
 FT DOMAIN 137 145 POLY-SER.  
 FT VARSPPLIC 631 643 EST -> GHS (in isoform 2).  
 FT VARSPPLIC 644 1327 /FTID=VSP\_004538.  
 FT VARSPPLIC 644 1327 Missing (in isoform 2).  
 FT MUTAGEN 1184 1184 /FTID=VSP\_004539.  
 FT MUTAGEN 1184 1184 H->A: LOSS OF ACTIVITY; WHEN ASSOCIATED  
 FT MUTAGEN 1291 1291 WITH A-1291.  
 FT MUTAGEN 1291 1291 E->A: LOSS OF ACTIVITY; WHEN ASSOCIATED  
 FT MUTAGEN 1291 1291 WITH A-1184.  
 SQ SEQUENCE 1327 AA; 142010 MW; E14DE985C710B957 CRC64;  
 Query Match 11.8%; Score 92; DB 1; Length 1327;  
 Best Local Similarity 30.4%; Pred. No. 5.2;  
 Matches 35; Conservative 15; Mismatches 57; Indels 8; Gaps 4;  
 QY 6 VAAFAAI-VVGSALAGVVPQWGGNGGNGGNSGDPSTLSIYQYGSANAALALQSDAR 64  
 DB 99 VAAAPVPAVSTSSAAGVAPNPAGSGSNSSSSPTSS-SSSSPSPGSSLAESPAA 157  
 QY 65 KSETTIT-----QSGYNGADYVGGADNSTIELTQNG--FRNATIDQWAKNSDI 113  
 DB 158 GVSSTAPLPGGAAGPGTGTVPVAVSGALRELLACRNGDVSRVRLVDAANVAKOM 212  
 RESULT 10  
 CSGB\_ECOLI  
 ID CSGB\_ECOLI STANDARD; PRT; 151 AA.  
 AC P39828;  
 DT 01-FEB-1995 (Rel. 31, Created)  
 DT 01-OCT-1996 (Rel. 34, Last sequence update)  
 DT 28-FEB-2003 (Rel. 41, Last annotation update)  
 DE Minor curlin subunit precursor.  
 GN CSGB OR B1041 OR Z1675 OR ECS1419.  
 OS Escherichia coli, and  
 OS Escherichia coli O157:H7.  
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;  
 OC Enterobacteriaceae; Escherichia.  
 OX NCBI\_TaxID=562, 83334;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=K12 / MC4100;  
 RX MEDLINE=9641466; PubMed=8817489;  
 RA Hammar M., Arqvist A., Bian Z., Olsen A., Normark S.;  
 RT "Expression of two csg operons is required for production of  
 RT fibronectin- and congo red-binding curli polymers in Escherichia coli  
 RT K-12.";  
 RL Mol. Microbiol. 18:661-670(1995).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=K12 / MG1655;  
 RX MEDLINE=97426617; PubMed=9278503;  
 RA Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,  
 RA Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.P.,  
 RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,  
 RA Mau B., Shao Y.;  
 RA "The complete genome sequence of Escherichia coli K-12.";  
 RL Science 277:1453-1474(1997).  
 RN [3]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=K12;  
 RX MEDLINE=97061202; PubMed=8905232;  
 RA Oshima T., Aiba H., Baba T., Fujita K., Hayashi K., Honjo A.,  
 RA Ikemoto K., Inada T., Itoh T., Kajihara M., Kanai K., Kashimoto K.,  
 RA Kimura S., Kitagawa M., Makino K., Masuda S., Miki T., Mizobuchi K.,  
 RA Mori H., Motomura K., Nakamura Y., Nishimoto H., Nishio Y., Saito N.,  
 RA Sampei G., Seki Y., Tagami H., Takemoto K., Wada C., Yamamoto Y.,  
 RA Yano M., Horiuchi T.;  
 RT "A 718-kb DNA sequence of the Escherichia coli K-12 genome

corresponding to the 12.7-28.0 min region on the linkage map.";

```

DB 89 NLAVIDQAGSANDASISQAGYNTAMIIQKSGNKANITQYGTOKTAIVV 138
RESULT 11
VG38_BPT2
ID VG38_BPT2 STANDARD; PRT; 262 AA.
AC P07875;
DT 01-AUG-1988 (Rel. 08, Created)
DT 01-AUG-1988 (Rel. 08, Last sequence update)
DT 01-AUG-1990 (Rel. 15, Last annotation update)
DE Receptor recognizing protein (Protein Gp38).
GN 38.
OS Bacteriophage T2.
OC Viruses; dsDNA viruses, no RNA stage; Caudovirales; Myoviridae;
OC T4-like viruses.
OX NCBI_TaxID=10664;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=87283911; PubMed=3302276;
RT "DNA sequence of genes 38 encoding a receptor-recognizing protein of
RT bacteriophages T2, K3 and of K3 host range mutants.";
RL J. Mol. Biol. 194:31-39 (1987).
CC -!- FUNCTION: VG38 is at the tip of the long tail fibers and serves as
CC the phage recognition site for the cellular receptor.
CC -!- MISCELLANEOUS: THIS PHAGE USE OUTER MEMBRANE PROTEINS OMPF AND TTR
CC AS RECEPTORS.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; X05312; CAA28935.1; -
CC PIR; S00275; S00275.
CC InterPro; IPR007932; Tail_fibre_GP38.
CC Pfam; PF05268; GP38; 1.
CC Fiber protein; Phage recognition.
KW
SQ SEQUENCE 262 AA; 25801 MW; 0567366918F6C745 CRC64;
Query Match 11.5%; Score 89.5; DB 1; Length 262;
Best Local Similarity 34.1%; Pred. No. 1.3;
Matches 30; Conservative 8; Mismatches 33; Indels 17; Gaps 4;
QY 27 GGGGHHNGGNSGPDSTLSIYQYGSANAALALQSDARKSETTITQSGYNGADYGGQAD 86
DB 175 GGGGRRPFGVGKIGSDSILS-----GSNASL---TDAGTGGTTF-QYGAGNGNVGAGG 225
QY 87 NSTIELTQNGFRNNATIDQWNAKNSDIT 114
DB 226 -----RGWKNVYTSSEGAAGAATV 245
RESULT 12
MSA2_PLAF2
ID MSA2_PLAF2 STANDARD; PRT; 347 AA.
AC Q03646;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 01-OCT-1996 (Rel. 34, Last annotation update)
DE Merozoite surface antigen 2 precursor (MSA-2).
GN MSA2.
OS Plasmodium falciparum (isolate Nig32 / Nigeria).
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_TaxID=70150;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=91156685; PubMed=2000383;
RA Smythe J.A.; Coppel R.L.; Day K.P.; Martin R.K.; Oduola A.M.J.;
RA Kemp D.J.; Anders R.F.;

```

```

RT corresponding to the 12.7-28.0 min region on the linkage map.";
RL DNA Res. 3:137-155 (1996).
RN [4]
RP SEQUENCE FROM N.A.
RC STRAIN=O157:H7 / EDL933 / ATCC 700927;
RX MEDLINE=21074935; PubMed=11206551;
RA Perna N.T.; Plunkett G. III; Burland V.; Mau B.; Glasner J.D.;
RA Rose D.J.; Mayhew G.F.; Evans P.S.; Gregor J.; Kirkpatrick H.A.;
RA Posfai G.; Hackett J.; Klink S.; Boutin A.; Shao Y.; Miller L.;
RA Grobeck E.J.; Davis N.W.; Lim A.; Dimallanta E.T.; Potanousis K.;
RA Apodaca J.; Anantharaman T.S.; Lin J.; Yen G.; Schwartz D.C.;
RA Welch R.A.; Blattner F.R.;
RT "Genome sequence of enterohaemorrhagic Escherichia coli O157:H7.";
RL Nature 409:529-533 (2001).
RN [5]
RP SEQUENCE FROM N.A.
RC STRAIN=O157:H7 / RIMD 0509952;
RX MEDLINE=21156231; PubMed=11256796;
RA Hayashi T.; Makino K.; Ohnishi M.; Kurokawa K.; Ishii K.; Yokoyama K.;
RA Han C.-G.; Ohtsubo E.; Nakayama K.; Murata T.; Tanaka M.; Tobe T.;
RA Iida T.; Takami H.; Honda T.; Sasaki C.; Ogasawara N.; Yasunaga T.;
RA Kuhara S.; Shiba T.; Hattori M.; Shinagawa H.;
RT "Complete genome sequence of enterohemorrhagic Escherichia coli
RT O157:H7 and genomic comparison with a laboratory strain K-12.";
RL DNA Res. 8:11-22 (2001).
RN [6]
RP SEQUENCE OF 1-21 FROM N.A.
RC STRAIN=K12;
RX MEDLINE=95157246; PubMed=7854117;
RA Arqvist A.; Olsen A.; Normark S.;
RT "Sigma S-dependent growth-phase induction of the csgBA promoter in
RT Escherichia coli can be achieved in vivo by sigma 70 in the absence
RT of the nucleoid-associated protein H-NS.";
RL Mol. Microbiol. 13:1021-1032 (1994).
CC -!- FUNCTION: CURLIN IS THE STRUCTURAL SUBUNIT OF THE CURLI. CURLI ARE
CC COILED SURFACE STRUCTURES THAT ASSEMBLE PREFERENTIALLY AT GROWTH
CC TEMPERATURES BELOW 37 DEGREES CELSIUS. CURLI CAN BIND TO
CC FIBRONECTIN. THE MINOR SUBUNIT IS THE NUCLEATION COMPONENT OF
CC CURLIN MONOMERS.
CC -!- SIMILARITY: BELONGS TO THE CSGA/CSGB FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; X90754; CAA62281.1; -
CC EMBL; AB000205; AAC74125.1; -
CC EMBL; D90741; BAA35831.1; -
CC EMBL; AE005315; AAG55787.1; -
CC EMBL; AP002554; BAB34842.1; -
CC PIR; C90806; C90806.
CC PIR; G85665; G85665.
CC PIR; S70787; S70787.
CC EcoGene; EGI2621; csgB.
KW Fimbria; Signal; Complete proteome.
FT SIGNAL 1 21 POTENTIAL.
FT CHAIN 22 151 MINOR CURLIN SUBUNIT.
SQ SEQUENCE 151 AA; 15882 MW; B18D266B964014B8 CRC64;
Query Match 11.7%; Score 91.5; DB 1; Length 151;
Best Local Similarity 29.1%; Pred. No. 0.48;
Matches 32; Conservative 17; Mismatches 54; Indels 7; Gaps 4;
QY 9 FAATVSSGAL--AGVTPQWGGGHHNGGNSGPDSTLSIY-QYGSANAALALQSDARK 65
DB 33 FAVNELSKSFTNQAAII---GQAGTNNASQLRQGGSKLLAVAGGSSNRA-KIDQTGY 88
QY 66 SETTITQSGYNGADYGGQADNSTIELTQNGFRNNATIDQWNAKNSDITV 115

```



Gordon S.V., Biglmeier K., Gas S., Barry C.E. III, Tekala F.,  
 Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,  
 Davies R., Devlin K., Feltwell T., Gentles S., Hamlin N., Holroyd S.,  
 Hornsby T., Jagels K., Krogh A., McLean J., Moule S., Murphy L.,  
 Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,  
 Rutter S., Seeger K., Skelton S., Squares S., Squares R.,  
 Sulston J.E., Taylor K., Whitehead S., Barrell B.G.;  
 "Deciphering the biology of Mycobacterium tuberculosis from the  
 complete genome sequence.";  
 Nature 393:537-544 (1998).  
 [2]  
 SEQUENCE FROM N.A.  
 SPECIES=M.tuberculosis; STRAIN=CDC 1551 / Oshkosh;  
 MEDLINE=22206454; PubMed=12218036;  
 RA Fleischmann R.D., Alland D., Eisen J.A., Carpenter L., White O.,  
 RA Peterson J.F., DeBoy R., Dodson R., Gwinn M., Haft D., Hickey E.,  
 RA Kolonay J.F., Nelson W.C., Umayam L.A., Ermolaeva M., Salzberg S.L.,  
 RA Delcher A., Utterback T., Weidman J., Khouri H., Gill J., Mikula A.,  
 RA Bishai W., Jacobs W.R. Jr., Venter J.C., Fraser C.M.;  
 "Whole-genome comparison of Mycobacterium tuberculosis clinical and  
 laboratory strains.";  
 J. Bacteriol. 184:5479-5490 (2002).  
 [3]  
 SEQUENCE FROM N.A.  
 SPECIES=M.bovis; STRAIN=AF2122/97;  
 MEDLINE=22709107; PubMed=12788972;  
 RA Garnier T., Eiglmeier K., Camus J.-C., Medina N., Mansoor H.,  
 RA Pryor M., Duthoit S., Grondin S., Lacroix C., Monsemp C., Simon S.,  
 RA Harris B., Atkin R., Doggett J., Mayes R., Keating L., Wheeler P.R.,  
 RA Parkhill J., Barrell B.G., Cole S.B., Gordon S.V., Hewinson R.G.;  
 "The complete genome sequence of Mycobacterium bovis.";  
 Proc. Natl. Acad. Sci. U.S.A. 100:7877-7882 (2003).  
 CC -1- SIMILARITY: BELONGS TO THE MYCOBACTERIAL PE FAMILY. PGRS  
 CC SUBFAMILY.  
 CC  
 CC -1- CAUTION: Ref.1 sequence differs from that shown due to a  
 frameshift in position 59. Ref.1 sequence has been checked by  
 authors in Ref.1 and they report that no errors have been found.  
 -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 the European Bioinformatics Institute. There are no restrictions on its  
 use by non-profit institutions as long as its content is in no way  
 modified and this statement is not removed. Usage by and for commercial  
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 or send an email to [license@sib-sib.ch](mailto:license@sib-sib.ch)).  
 -----  
 CC EMBL: Z73966; CAA98228.1; ALT FRAME.  
 CC EMBL: AX007065; AAK46440.1; -  
 CC EMBL: BE248341; CAD96978.1; -  
 CC TIGR: MT2159; -  
 CC TubercuList; RV2098c; -  
 CC InterPro; IPR000084; PE\_region.  
 CC Pfam; PF00934; PE; 1.  
 CC ProDom; PD001223; PE\_region; 1.  
 CC Hypothetical protein; Complete proteome.  
 CC CONFLICT 312 312 G -> GG (IN REF. 1).  
 CC SEQUENCE 491 AA; 41979 MW; 12C9630C59CA0C13 CRC64;  
 Query Match 11.3%; Score 88; DB 1; Length 491;  
 Best Local Similarity 26.8%; Pred.No. 3.6;  
 Matches 30; Conservative 13; Mismatches 39; Indels 30; Gaps 5;  
 QY 20 AGVVPVQGGGNN-----HNGGNSGSPDSTLSIYQGSNAALQSDARKSETTITQSG 74  
 Db 222 AGLIHGGAGGNGDGHGGSGHAGSGGSGGFGFGAGGLI----- 264  
 QY 75 YNGNADYVGQAD--NSTIEITQNGFRNATIDQWNAKNSD---ITVQYQVGN 121  
 Db 265 YNGGGAAGSGGNGDAGTGVSSDGF---AGLGGSGRGGDAGLIGVGGGGGN 313  
 RESULT 14  
 OMB2 NEIMB

RESULT 15

PER DROWI STANDARD; PRT; 1093 AA.

ID Q03297; O18422; O18422; P91721; P91722;

DT 01-OCT-1993 (Rel. 27, Created)

DT 15-JUL-1998 (Rel. 36, Last sequence update)

DT 16-OCT-2001 (Rel. 40, Last annotation update)

DE Period circadian protein (fragment).

GN PER.

OS Drosophila willistoni (Fruit fly).

OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;

OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;

OC Ephydroidea; Drosophilidae; Drosophila.

OX NCBI\_TaxID=7260;

RA [1]

RN SEQUENCE FROM N.A.

RP STRAIN=Various strains;

RC MEDLINE=97357421; PubMed=9214747;

RA Gleason J.M., Powell J.R.;

RT "Interspecific and intraspecific comparisons of the period locus in the Drosophila willistoni sibling species.";

RT Mol. Biol. Evol. 14:741-753(1997).

RL [2]

RN SEQUENCE OF 579-646 FROM N.A.

RP MEDLINE=93196482; PubMed=8450754;

RA Peixoto A.A., Campos S., Costa R.H., Kyriacou C.P.;

RT "Molecular evolution of a repetitive region within the per gene of Drosophila.";

RL Mol. Biol. Evol. 10:127-139(1993).

CC -!- FUNCTION: Essential for biological clock functions. Determines the period length of circadian and ultradian rhythms; an increase in PER dosage leads to shortened circadian rhythms and a decrease leads to lengthened circadian rhythms. Essential for the circadian rhythmicity of locomotor activity, eclosion behavior, and for the rhythmic component of the male courtship song that originates in the thoracic nervous system. The biological cycle depends on the rhythmic formation and nuclear localization of the TIM-PER complex. Light induces the degradation of TIM, which promotes elimination of PER. Nuclear activity of the heterodimer coordinatively regulates PER and TIM transcription through a negative feedback loop. Behaves as a negative element in circadian transcriptional loop. Does not appear to bind DNA, suggesting indirect transcriptional inhibition (by similarity).

CC -!- SUBUNIT: Forms heterodimer with timeless (TIM); the complex then translocates into the nucleus (by similarity).

CC -!- SUBCELLULAR LOCATION: Nuclear at specific periods of the day. First accumulates in the perinuclear region about one hour before translocation into the nucleus. Interaction with TIM is required for nuclear localization (by similarity).

CC -!- PTM: PHOSPHORYLATED WITH A CIRCADIAN RHYTHMICITY, PROBABLY BY THE DOUBLE-TIME PROTEIN (DBT). PHOSPHORYLATION COULD BE IMPLICATED IN THE STABILITY OF PER MONOMER AND IN THE FORMATION OF HETERODIMER PER-TIM (BY SIMILARITY).

CC -!- MISCELLANEOUS: THE SEQUENCE SHOWN IS THAT OF STRAIN ATLIXCO.

CC -!- SIMILARITY: Contains 2 PAS (PER-ARNT-SIM) dimerization domains.

CC -!- SIMILARITY: Contains 1 PAS-associated C-terminal (PAC) domain.

CC -----

CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC -----

CC EMBL; U51055; AAB41360.1; -

CC EMBL; U51056; AAB41361.1; -

CC EMBL; U51057; AAB41362.1; -

CC EMBL; U51058; AAB41363.1; -

CC EMBL; U51059; AAB41364.1; -

CC EMBL; U51060; AAB41365.1; -

CC EMBL; U51061; AAB41366.1; -

CC EMBL; U51062; AAB41367.1; -

```
DR EMBL; U51063; AAB41368.1; --
DR EMBL; U51064; AAB41369.1; --
DR EMBL; U51065; AAB41370.1; --
DR EMBL; U51066; AAB41371.1; --
DR EMBL; U51067; AAB41372.1; --
DR EMBL; U51068; AAB41373.1; --
DR EMBL; U51069; AAB41374.1; --
DR EMBL; U51070; AAB41375.1; --
DR EMBL; U51071; AAB41376.1; --
DR EMBL; U51072; AAB41377.1; --
DR EMBL; L06342; AAA28765.1; --
DR FlyBase; FBgn0013161; Dm11\ber.
DR InterPro; IPR000014; PAS_domain.
DR Pfam; PF00989; PAS; 2.
DR SMART; SM00091; PAS; 2.
DR PROSITE; PS50112; PAS; 2.
KW Biological rhythms; Repeat; Nuclear protein; Phosphorylation;
KW Polymorphism.
FT NON_TER 1 1
FT DOMAIN <1 12 NUCLEAR LOCALIZATION SIGNAL (POTENTIAL).
FT DOMAIN 139 209 PAS 1.
FT DOMAIN 289 359 PAS 2.
FT DOMAIN 371 411 PAC.
FT DOMAIN 7 12 POLY-LYS.
FT DOMAIN 618 625 POLY-GLY.
FT DOMAIN 718 734 POLY-GLY.
FT DOMAIN 745 748 POLY-SER.
FT DOMAIN 759 770 POLY-GLY.
FT DOMAIN 885 888 POLY-ALA.
FT DOMAIN 911 917 POLY-ALA.
FT VARIANT 611 611 T -> A (IN STRAIN 0811.4).
FT VARIANT 617 617 S -> F (IN STRAIN 0811.4).
FT VARIANT 622 622 G -> V (IN STRAIN GUANA).
FT VARIANT 724 724 G -> A (IN STRAIN MANAUS 2).
FT VARIANT 726 726 G -> S (IN STRAIN SANTA MARIA).
FT VARIANT 729 734 MISSING (IN STRAIN PORTO ALEGRE 3).
FT VARIANT 730 734 MISSING (IN STRAIN PORTO ALEGRE 4).
FT VARIANT 731 734 MISSING (IN STRAINS MANAUS 1 AND MANAUS 3).
FT VARIANT 732 734 MISSING (IN STRAINS LIMA B, L'HABITATUE AND CANO MORA).
FT VARIANT 733 733 G -> V (IN STRAIN PORTO ALEGRE 4).
FT VARIANT 733 734 MISSING (IN STRAINS GUADELOUPE, MANAUS 2, PORTO ALEGRE 2, PORTO ALEGRE 1 AND GUANA).
FT VARIANT 734 734 MISSING (IN STRAINS MANAUS 4, PORTO ALEGRE 1 AND PORTO ALEGRE 2).
FT VARIANT 747 747 S -> A (IN STRAINS GUADELOUPE AND GUANA).
FT VARIANT 764 766 MISSING (IN STRAIN MANAUS 3).
FT VARIANT 886 886 A -> T (IN STRAIN 0811.4).
FT NON_TER 1093 1093
SQ SEQUENCE 1093 AA; 115896 MW; AB6DE050267EC187 CRC64;

Query Match 11.0%; Score 86; DB 1; Length 1093;
Best Local Similarity 27.2%; Pred.No. 13;
Matches 25; Conservative 6; Mismatches 45; Indels 16; Gaps 2;

QY 27 GCGGNHNGGNSGPDSTLSIYQGSANAALQSDARKSETITQSGYNGADYQGAD 86
Db 721 GGGGGGGGGGGGGGLPLFLDVHTSS-----SSQNKGTGVAAGGAGGGVGGGG-- 770
QY 87 NSTIELTQGFENNATIDOWNAKNSDITVGY 118
Db 771 -----SCSLGNGNGVSGNGNSQFSTNQY 796
```

Search completed: August 2, 2004, 14:49:27  
Job time : 6.3 secs

This Page Blank (uspto)

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: August 2, 2004, 14:39:17 ; Search time 29.7 Seconds

(without alignments)  
1604.150 Million cell updates/sec

Title: US-09-543-407-12

Perfect score: 779

Sequence: 1 MKLLKVAFAAIVVSGSALA.....VTRVTHEMAHANNATANQY 151

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1017041 seqs, 315518202 residues

Total number of hits satisfying chosen parameters: 1017041

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

SPTREMBL\_25:  
1: sp\_archaea:  
2: sp\_bacteria:  
3: sp\_fungi:  
4: sp\_human:  
5: sp\_invertebrate:  
6: sp\_mammal:  
7: sp\_mhc:  
8: sp\_organelle:  
9: sp\_phage:  
10: sp\_plant:  
11: sp\_rodent:  
12: sp\_virus:  
13: sp\_vertebrate:  
14: sp\_unclassified:  
15: sp\_virus:  
16: sp\_bacteriopl:  
17: sp\_archaeap:

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	686	88.1	152	2	Q33802
2	591.5	75.9	150	2	Q7X243
3	537	68.9	149	2	Q7X240
4	495.5	63.6	152	16	Q8CW63
5	427.5	54.9	150	2	Q7X237
6	385	49.4	76	2	Q54069
7	122	15.7	29	2	Q9S3J5
8	113.5	14.6	3501	16	Q8Y106
9	113.5	14.6	3552	16	Q8XSD6
10	108.5	13.9	191	10	Q7XDR3
11	107	13.7	502	16	Q8EIH4
12	106	13.6	1748	5	Q94821
13	105	13.5	2035	2	Q9XCJ4
14	105	13.5	2039	16	Q8ZN57
15	104.5	13.4	1286	2	Q841Y5
16	104	13.4	2174	16	Q92U08

17	103	13.2	139	16	Q8EIH3
18	102.5	13.2	624	3	Q8NIV1
19	102	13.1	1422	16	Q3EFU3
20	100.5	12.9	152	2	Q7X241
21	100	12.8	179	2	Q33801
22	99.5	12.8	151	2	Q7X244
23	99	12.7	1765	16	Q7V8S5
24	98.5	12.6	1222	5	Q9W4F0
25	98.5	12.6	1222	5	Q8T8L9
26	98.5	12.6	1351	5	Q8SX56
27	98.5	12.6	1354	5	Q8MPN4
28	98.5	12.6	7716	16	Q7UWZ8
29	97.5	12.5	154	16	Q89J15
30	97.5	12.5	348	13	Q93397
31	97.5	12.5	739	2	Q9X687
32	97.5	12.5	3659	16	Q98LN6
33	97	12.5	1618	2	Q9KKB1
34	96.5	12.4	151	2	Q7X238
35	96	12.3	145	16	Q8U6N9
36	96	12.3	157	16	Q8HGO
37	95.5	12.3	130	16	Q89J14
38	95.5	12.3	453	5	Q9N6M8
39	95.5	12.3	1615	2	Q9KKA8
40	95	12.2	362	16	Q8EV84
41	95	12.2	1613	2	Q840U5
42	95	12.2	1618	2	Q9KKB4
43	94.5	12.1	160	16	Q8CW64
44	94.5	12.1	453	5	Q9NGF6
45	94.5	12.1	453	5	Q9NGF7

#### ALIGNMENTS

RESULT 1  
Q33802  
ID O33802 PRELIMINARY; PRT; 152 AA.  
AC O33802;  
DT 01-JAN-1998 (TRENBLrel. 05, Created)  
DT 01-JAN-1998 (TRENBLrel. 05, Last sequence update)  
DT 01-DEC-2001 (TRENBLrel. 19, Last annotation update)  
DE AgfA protein (Fragment).  
GN AGFA.  
OS Salmonella typhimurium.  
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;  
OC Enterobacteriaceae; Salmonella.  
OX NCBI\_TaxID=602;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=98053981; PubMed=9393832;  
RA Sukupolvi S.S., Lorentz R.G., Gordon J.I., Bian Z., Pfeifer J.D.,  
RA Normark S.J., Ren M.,  
RT "Expression of thin, aggregative fimbriae promotes interaction of  
RT Salmonella typhimurium SR-11 with mouse small intestinal epithelial  
RT cells."  
RL Infect. Immun. 65:5320-5325(1997).  
DR EMBL; AJ000514; CAA04151.1; --  
FT NON TER 152 152  
SQ SEQUENCE 152 AA; 15401 MW; 9DA7DADC2364B006 CRC64;  
Query Match 88.1%; Score 686; DB 2; Length 152;  
Best Local Similarity 88.7%; Pred. No. 4.1e-45;  
Matches 134; Conservative 5; Mismatches 12; Indels 0; Gaps 0;  
QY 1 MKLLKVAFAAIVVSGSALAIVVPGWGGGNGHNGSGSPDSTLSIYQVGSANAALALQ 60  
DB 1 MKLLKVAFAAIVVSGSALAIVVPGWGGGNGHNGSGSPDSTLSIYQVGSANAALALQ 60  
QY 61 SDARKSETTITQSGYNGADVGQCADNSTIELTQNGFRNATIDQWNAKNSDITVGYGG 120  
DB 61 SDARKSETTITQSGYNGADVGQCADNSTIELTQNGFRNATIDQWNAKNSDITVGYGG 120  
QY 121 NNALVNYDQLVTRVTHEMAHANNATANQY 151

```

Db      121  NNAALVNQTASDSSVWVQVGFNNAPANQY 151
||||||| : : : |||||
RESULT 2
Q7X243
ID Q7X243 PRELIMINARY; PRT; 150 AA.
AC Q7X243;
DT 01-OCT-2003 (TrEMBLrel. 25, Created)
DT 01-OCT-2003 (TrEMBLrel. 25, Last sequence update)
DE 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Curlin-csgA protein.
GN CSQA.
OS Citrobacter sp. Fec2.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Citrobacter.
OX NCBI_TaxID=213763;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Fec2;
RA Zogaj X., Bokranz W., Nimitz M., Romling U.;
RT "Production of Cellulose and Curli Fimbriae by Members of the Family
RT Enterobacteriaceae Isolated from the Human Gastrointestinal Tract.";
RL Infect. Immun. 72:4151-4158 (2003).
DR EMBL: AJ515700; CAD56672.1; -.
SQ SEQUENCE 150 AA; 15016 MW; 1D7141B8D6973DC6 CRC64;

Query Match 75.9%; Score 591.5; DB 2; Length 150;
Best Local Similarity 77.5%; Pred. No. 6.7e-38;
Matches 117; Conservative 14; Mismatches 19; Indels 1; Gaps 1;

QY      1  MKLLKVAFAAIVVSGSALAGVVPQWGGGNGHNGSSGPDSTLSIYQYGSANAALAQ 60
Db      1  MKLLQVAPAAIVVSGSALAGVVPQWGGGNGHNGSSGPDSTLSIYQYGSANAALAQ 59
||||||| : : : |||||
QY      61  SDARKSETTITQSGYNGADVGQGDNSTIELTQNGFRNNATIDQWNAKNSDITVQYGG 120
Db      60  SDARKSDTTIHQNGFGADVGQGSNDSTIDLTQNGFRNNATIDQWNGKNSDITVQYGG 119
||||||| : : : |||||
QY      121  NNAALVNVDQLVTRVVTHEMAHANNATANQY 151
Db      120  HNAALVNQTASDSSVWVQVGFNNAPANQY 150
||||||| : : : |||||

RESULT 3
Q7X240
ID Q7X240 PRELIMINARY; PRT; 149 AA.
AC Q7X240;
DT 01-OCT-2003 (TrEMBLrel. 25, Created)
DT 01-OCT-2003 (TrEMBLrel. 25, Last sequence update)
DE 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Curlin-csgA protein.
GN CSQA.
OS Citrobacter freundii.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Citrobacter.
OX NCBI_TaxID=546;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Fec4;
RA Zogaj X., Bokranz W., Nimitz M., Romling U.;
RT "Production of Cellulose and Curli Fimbriae by Members of the Family
RT Enterobacteriaceae Isolated from the Human Gastrointestinal Tract.";
RL Infect. Immun. 72:4151-4158 (2003).
DR EMBL: AJ515701; CAD56675.1; -.
SQ SEQUENCE 149 AA; 15200 MW; 946DD52017F648FD CRC64;

Query Match 68.9%; Score 537; DB 2; Length 149;
Best Local Similarity 71.5%; Pred. No. 9.7e-34;
Matches 108; Conservative 17; Mismatches 24; Indels 2; Gaps 1;

QY      1  MKLLKVAFAAIVVSGSALAGVVPQWGGGNGHNGSSGPDSTLSIYQYGSANAALAQ 60
||||||| : : : |||||

```

```

RT "Production of Cellulose and Curli Fimbriae by Members of the Family
RT Enterobacteriaceae Isolated from the Human Gastrointestinal Tract.";
RL Infect. Immun. 72:4151-4158(2003).
DR EMBL: AJ515702; CAD56678.1; -.
SQ SEQUENCE 150 AA; 15112 MW; 5D8BE2D872DF15F3 CRC64;

Query Match 54.9%; Score 427.5; DB 2; Length 150;
Best Local Similarity 58.3%; Pred. No. 2.3e-25;
Matches 88; Conservative 26; Mismatches 36; Indels 1; Gaps 1;

QY 1 MKLLKVAFAAIVVSGSALAGVVPQGGGHNHGGNSGPDSTLSIYQGSANALALQ 60
DQ :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 1 MKFIKVAALAAIVVSGSAMAGMIQ-CGWGHGHGGYGGPSTLNIYQNGGNSALALQ 59
QY 61 SDARKSETTTTQSGYNGADVGGQADNSTIELTQNGFRNNATIDQWAKNSDITVGGYGG 120
DQ :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 60 TDARNSVLNIQSQTGGGADVGGQSDSSINLTQNGFGNSATIDQWNSKDSVNVVSGYGG 119
QY 121 NNAALVNYDQIVTRVTHVMAHANNATANY 151
DQ :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 120 LAGALVDQTASNTVNTVTCIGFGNHATAHQY 150

RESULT 6
Q54069 PRELIMINARY; PRT; 76 AA.
AC Q54069;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE SEF17 fimbria (Fragment).
GN AGFA.
OS Salmonella enteritidis.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Salmonella.
OX NCBI_TaxID=592;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=SE30;
RA Cox J.M., Eglezos S., Woolcock J.B.;
RT "Virulence of Salmonella enteritidis in chickens correlates with
RT colony morphology and expression of SEF17 fimbriae.";
RL Submitted (APR-1996) to the EMBL/GenBank/DBJ databases.
DR EMBL: U53207; AAA98671.1; -.
FT NON_TER 1
FT NON_TER 76
SQ SEQUENCE 76 AA; 7704 MW; 2FD5411241A7BCB1 CRC64;

Query Match 49.4%; Score 385; DB 2; Length 76;
Best Local Similarity 98.7%; Pred. No. 1.9e-22;
Matches 75; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 30 GNHNGGNSGPDSTLSIYQGSANALALQSDARKSETTTTQSGYNGADVGGQADNST 89
DQ :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 1 GNHNGGNSGPDSTLSIYQGSANALALQSDARKSETTTTQSGYNGADVGGQADNST 60
QY 90 IELTQNGFRNNATIDQ 105
DQ :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 61 IELTQNGFRNNATIDQ 76

RESULT 7
Q9S3J5 PRELIMINARY; PRT; 29 AA.
AC Q9S3J5;
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT 01-MAR-2001 (TrEMBLrel. 16, Last annotation update)
DE Curlin subunit monomer (fragment).
GN CSGA.
OS Escherichia coli.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Escherichia.

```

```

OX NCBI_TaxID=562;
RN [1]
RP SEQUENCE FROM N.A.
RC TRANSPOSON-insertion sequence IS1;
RX MEDLINE=99314153; PubMed=10386375;
RA La Ragione R.M., Collighan R.J., Woodward M.J.;
RT "Non-cultivation of Escherichia coli O78:K80 isolates associated with
RT IS1 inserti on in csqB and reduced persistence in poultry infection.";
RL FEMS Microbiol. Lett. 175:247-253(1999).
DR EMBL: AJ131756; CAB45380.1; -.
FT NON_TER 29
FT NON_TER 29
SQ SEQUENCE 29 AA; 2789 MW; E290DFC07ABBB243 CRC64;

Query Match 15.7%; Score 122; DB 2; Length 29;
Best Local Similarity 89.7%; Pred. No. 0.0078;
Matches 26; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 MKLLKVAFAAIVVSGSALAGVVPQGGG 29
DQ :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 1 MKLLKVAATAAIVFSGSALAGVVPQYGGG 29

RESULT 8
Q8Y106 PRELIMINARY; PRT; 3501 AA.
AC Q8Y106;
DT 01-MAR-2002 (TrEMBLrel. 20, Created)
DT 01-MAR-2002 (TrEMBLrel. 20, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Probable hemagglutinin-related protein.
GN RSC0887 OR RS06116.
OS Ralstonia solanacearum (Pseudomonas solanacearum).
OC Bacteria; Proteobacteria; Betaproteobacteria; Burkholderiales;
OC Burkholderiaceae; Ralstonia.
OX NCBI_TaxID=305;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CM11000;
RX MEDLINE=21681879; PubMed=11823852;
RA Salanoubat M., Genin S., Artiguenave F., Gouzy J., Mangelot S.,
RA Arlat M., Billault A., Brottier P., Camus J.C., Cattolico L.,
RA Chandler M., Choisme N., Claudel-Renard C., Cunnac S., Demange N.,
RA Gaspin C., Lavie M., Moisan A., Robert C., Saurin W., Schiex T.,
RA Siguer P., Thebault P., Whalen M., Wincker P., Levy M.,
RA Weissenbach J., Boucher C.A.;
RT "Genome sequence of the plant pathogen Ralstonia solanacearum.";
RL Nature 415:497-502(2002).
DR EMBL: AL646061; CAD14589.1; -.
DR GO: GO:0004519; F:endonuclease activity; IEA.
DR GO: GO:0003676; F:nucleic acid binding; IEA.
DR InterPro: IPR001604; Endonuclease.
DR InterPro: IPR008619; F11_haemagg.
DR InterPro: IPR008638; Haemagg_act.
DR Pfam: PF05594; F11_haemagg; 20.
DR Pfam: PF05860; Haemagg_act; 1.
DR PROSITE: PS01070; NUCLEASE_NON_SPEIC; 1.
KW Complete proteome.
SQ SEQUENCE 3501 AA; 348421 MW; 290B41C99018A107 CRC64;

Query Match 14.6%; Score 113.5; DB 16; Length 3501;
Best Local Similarity 28.8%; Pred. NO. 8.3;
Matches 36; Conservative 19; Mismatches 43; Indels 27; Gaps 4;

QY 15 SGSALAGVVPQGGGNGGG-NSSGPDSTLSIYQGSANAA----- 56
DQ :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 2431 SGSHFTAGPSWGLGRNVGGPNSSG----VGLAPYGSASNAAGNSRQNASVVIG 2486
QY 57 LALQSDARKSETTTTQSGYNGADVGGQADNSTIELTQNGFRNNATIDQWAKNSDITV 116
DQ :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 2487 KSVQVQARTGDTVTSVSGSISALSDVLLAKQKQDIVAGNDTSRHH-----HSDRTIG 2541
QY 117 QYGGN 121
DQ :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:

```





QY 122 N 122  
Db 350 N 350

RESULT 12

Q94821 Q94821 PRELIMINARY; PRT; 1748 AA.  
AC Q94821; P92146; P92145; P92144; P92143; P92142; P92141; Q94820;  
DT 01-FEB-1997 (TrEMBLrel. 02, Created)  
DT 01-FEB-1997 (TrEMBLrel. 02, Last sequence update)  
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)  
DE CNUB protein.  
GN CNUB.  
OS Tetrahymena thermophila.  
OC Eukaryota; Alveolata; Ciliophora; Oligohymenophorea; Hymenostomatida;  
OC Tetrahymenina; Tetrahymena.  
OC NCBI\_TaxID=5911;  
OX [1]  
RN [1]  
RP MEDLINE=88189811; PubMed=3357771;  
RX Martindale D.W., Taylor F.M.;  
RA "Multiple introns in a conjugation-specific gene from Tetrahymena thermophila";  
RT Nucleic Acids Res. 16:2189-2201(1988).  
RL [2]  
RN [2]  
RP MEDLINE=94051569; PubMed=8233798;  
RX Taylor F.M., Martindale D.W.;  
RA "Retroviral-type zinc fingers and glycine-rich repeats in a protein encoded by cnjB, a Tetrahymena gene active during meiosis";  
RT Nucleic Acids Res. 21:4610-4614(1993).  
RL ENBL; X06462; CAB37323.1; -;  
DR ENBL; L03710; AAC37171.1; -;  
DR PIR; S42136; S42136.  
DR HSSP; P05888; 1A4F.  
DR GO; GO:0003676; F nucleic acid binding; IEA.  
DR InterPro; IPR001878; Znf\_CCHC.  
DR Pfam; PF00098; zif\_CCHC; 7.  
DR PRINTS; PR00939; C2HCZNFINGER.  
DR SMART; SM00343; ZNF\_C2HC; 7.  
DR PROSITE; PS0158; ZF\_CCHC; 7.  
FT CONFLICT 251 251 M -> I (IN REF. 1).  
FT CONFLICT 256 256 I -> N (IN REF. 1).  
SQ SEQUENCE 1748 AA; 199624 MW; 0B03F210104008A3 CRC64;

Query Match 13.6%; Score 106; DB 5; Length 1748;  
Best Local Similarity 32.1%; Pred. No. 14;  
Matches 35; Conservative 13; Mismatches 33; Indels 28; Gaps 5;

QY 25 QWGGGNGHGG---GNSSGPDSTLSTLYQVGSANALALQSDARKSETTTT---QSGYGN 77  
Db 1640 QFGGGGNSGGGSGWGTSGSDWN-----CQSNVQESTTTSSGGWSSGSGN 1685

QY 78 GADVGGGADNSTLTQNGFRNNATIDOWNAKNSDITVGC--YGGNAA 124  
Db 1686 QTGGGWSNDN-----QQQNTGGGGSSNSQNTNNESSWGNQA 1729

RESULT 13

Q9XCJ4 Q9XCJ4 PRELIMINARY; PRT; 2035 AA.  
ID Q9XCJ4  
AC Q9XCJ4;  
DT 01-NOV-1999 (TrEMBLrel. 12, Created)  
DT 01-NOV-1999 (TrEMBLrel. 12, Last sequence update)  
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)  
DE ShdA.  
GN SHdA.  
OS Salmonella typhimurium.  
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;  
OC Enterobacteriaceae; Salmonella.  
OX NCBI\_TaxID=602;  
RN [1]

RP SEQUENCE FROM N.A.  
RC STRAIN=ATCC14028;  
RA Kingsley R.A., van Amsterdam K., Bauml A.J.;  
RT "The presence of a pathogenicity island specific to Salmonella enterica subspecies I correlates with adaptation to warm blooded animals";  
RL Submitted (JUN-1999) to the EMBL/GenBank/DBJ databases.  
RN [2]  
RP SEQUENCE FROM N.A.  
RC STRAIN=ATCC14028;  
RA Kingsley R.A., van Amsterdam K., Edwards E.W., Hargis B.M., Bauml A.J.;  
RT "Complete sequence of the xseA-hisS intergenic region of the S. enterica serotype Typhimurium genome and its distribution within the genus Salmonella";  
RL Submitted (JUN-1999) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AF140550; AAD25110.2; -;  
DR InterPro; IPR006315; Autotransport.  
DR InterPro; IPR005546; Autotransporter.  
DR InterPro; IPR004899; Pertactin.  
DR InterPro; IPR002173; PfkB.  
DR Pfam; PF03797; Autotransporter; 1.  
DR Pfam; PF03212; Pertactin; 1.  
DR TIGRFAMs; TIGR01414; autotrans barl; 3.  
DR PROSITE; PSC0584; PFKB KINASES 2; 2.  
SQ SEQUENCE 2035 AA; 207032 MW; 295DB82FFCA84FAB CRC64;

Query Match 13.5%; Score 105; DB 2; Length 2035;  
Best Local Similarity 26.3%; Pred. No. 20;  
Matches 54; Conservative 20; Mismatches 61; Indels 70; Gaps 11;

QY 10 AAVVSGSALAG-----VVPQWGGGNGHGG---NSGPD----- 42  
Db 90 AALVVGATVGTGVLVETSGGADDPGKYSNAISLDHYAILSLTDAKI 149

QY 43 STLSTYQVGSANA-----ALALQSDARKS-ETTITQSGYNGADVG 82  
Db 150 TTTGIYTGISAGDSTLRLTDLTIDNGFVMTLTGSEATLDTIVEAANSSAQVQ 209

QY 83 QGA-----DNSTIELTO---NGFRNNATIDOWNAKN-SDITVGVG-----GNNAALVN 127  
Db 210 QGSTLVLDGSTITLQAGQINNVAGNTATDEGTLNLSDSVSSAGTMTSTIQINKALN 269

QY 128 YDQLVTRVVTHEMA-----HANNAT 147  
Db 270 ---LTNATITHTNAGAAVQANNAT 291

RESULT 14

Q8ZN57 Q8ZN57 PRELIMINARY; PRT; 2039 AA.  
ID Q8ZN57  
AC Q8ZN57;  
DT 01-MAR-2002 (TrEMBLrel. 20, Created)  
DT 01-MAR-2002 (TrEMBLrel. 20, Last sequence update)  
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)  
DE Similar to the C-terminal region of AIDA, IcsA, subspecies I specific, DE Peyer's patch colonization and shedding factor.  
GN SHdA OR STM2513.  
OS Salmonella typhimurium.  
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;  
OC Enterobacteriaceae; Salmonella.  
OX NCBI\_TaxID=602;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=LT2 / SGSC1412 / ATCC 700720;  
RX MEDLINE=21534948; PubMed=11677609;  
RA McClelland M., Sanderson K.E., Spieth J., Clifton S.W., Latreille P., Courtney L., Porwollik S., Ali J., Dante M., Du F., Hou S., Layman D., Leonard S., Nguyen C., Scott K., Holmes A., Grewal N., Mulvaney E., Ryan E., Sun H., Florea L., Miller W., Stoneking T., Nhan M., Waterston R., Wilson R.K.;  
RT "Complete genome sequence of Salmonella enterica serovar Typhimurium LT2.";

```

RL Nature 413:852-856(2001).
DR EMBL; AF008813; AAL21407.1; -.
DR InterPro; IPR006315; Autotransport.
DR InterPro; IPR005546; Autotransporter.
DR InterPro; IPR004899; Pertactin.
DR InterPro; IPR002173; PFKB.
DR Pfam; PF03797; Autotransporter; 1.
DR Pfam; PF03212; Pertactin; 1.
DR TIGRfams; TIGR01414; Autotrans barl; 3.
DR PROSITE; PS00584; PFKB_KINASES_2; 2.
KW Complete proteome.
SQ
SEQUENCE 2039 AA; 207127 MW; 894E41F8F29339EA CRC64;

Query Match
Best Local Similarity 13.5%; Score 105; DB 16; Length 2039;
Matches 54; Conservative 20; Mismatches 51; Indels 70; Gaps 11;

QY 10 AAVVSGSALAG-----VVPQWGGGNGNGG-----NSSQPD----- 42
Db 94 AALYVSGVATVGMQPTVTGRLVETSGGADPDGKYVSNALSDHYAILELTDKI 153
QY 43 STLSTYQYGSANA-----ALALOSDARKS-ETTITQSGYNGADV 82
Db 154 TTGTYTGISAAAGSTLRLTSDTLTIDGNGVWTLTSGEALDGTIVEANSSAQVQ 213
QY 83 QGA-----DNSTIELTQ--NGFRNATIOWNAKN-SDITVQYQ-----GNNAAVYN 127
Db 214 QGSLTNVLDDGSTITTLAQQINWAGNTATDEGSTLNLSDSVSSAGTMTSIQGTNKAALN 273
QY 128 YDQLVTRVVTHEMA-----HANNAT 147
Db 274 ---LTNATIHTNAGAAVQANNAT 295

RESULT 15
Q841Y5 PRELIMINARY; PRT; 1286 AA.
AC Q841Y5;
CT 01-JUN-2003 (TRENBLrel. 24, Created)
DT 01-JUN-2003 (TRENBLrel. 24, Last sequence update)
DE Putative high-molecular-weight surface-exposed protein Cf0009.
OS Campylobacter fetus.
OC Bacteria; Proteobacteria; Epsilonproteobacteria; Campylobacteriales;
OC Campylobacteraceae; Campylobacter.
OX NCBI_TaxID=196;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=23D;
RX MEDLINE=90354448; PubMed=2387868;
RA Blaser M.J., Gotschlich E.C.;
RT "Surface array protein of Campylobacter fetus. Cloning and gene
structure."
RL J. Biol. Chem. 265:14529-14535(1990).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=23D;
RX MEDLINE=91035477; PubMed=2229082;
RA Blaser M.J., Gotschlich E.C.;
RT "Surface array protein of Campylobacter fetus. Cloning and gene
structure."
RL J. Biol. Chem. 265:19372-19372(1990).
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=23D;
RX MEDLINE=92394895; PubMed=1522068;
RA Tummuru M.K., Blaser M.J.;
RT "Characterization of the Campylobacter fetus sapA promoter: evidence
that the sapA promoter is deleted in spontaneous mutant strains."
RL J. Bacteriol. 174:5916-5922(1992).
RN [4]
RP SEQUENCE FROM N.A.
RC STRAIN=23D;

```

```

RX MEDLINE=93348254; PubMed=8346244;
RA Tummuru M.K., Blaser M.J.;
RT "Rearrangement of sapA homologs with conserved and variable regions in
Campylobacter fetus."
RL Proc. Natl. Acad. Sci. U.S.A. 90:7265-7269(1993).
RN [5]
RP SEQUENCE FROM N.A.
RC STRAIN=23D;
RX MEDLINE=95204338; PubMed=7896695;
RA Dworkin J., Tummuru M.K., Blaser M.J.;
RT "A lipopolysaccharide-binding domain of the Campylobacter fetus S-
layer protein resides within the conserved N terminus of a family of
silent and divergent homologs."
RL J. Bacteriol. 177:1734-1741(1995).
RN [6]
RP SEQUENCE FROM N.A.
RC STRAIN=23D;
RX MEDLINE=99069317; PubMed=9851986;
RA Thompson S.A., Sheeh O.L., Ray K.C., Beins M.H., Jorgensen J.P.,
Blaser M.J.;
RT "Campylobacter fetus surface layer proteins are transported by a type
I secretion system."
RL J. Bacteriol. 180:6450-6458(1998).
RN [7]
RP SEQUENCE FROM N.A.
RC STRAIN=23D;
RX PubMed=12694614;
RA Tu Z.C., Wassenaar T.M., Thompson S.A., Blaser M.J.;
RT "Structure and genotypic plasticity of the Campylobacter fetus sap
locus."
RL Mol. Microbiol. 48:685-698(2003).
DR EMBL; AY211269; AAC64216.1; -.
DR InterPro; IPR005546; Autotransporter.
DR Pfam; PF03797; Autotransporter; 1.
DR SEQUENCE 1286 AA; 134079 MW; A1FF9CB34158789 CRC64;

Query Match
Best Local Similarity 13.4%; Score 104.5; DB 2; Length 1286;
Matches 44; Conservative 26; Mismatches 57; Indels 31; Gaps 9;

QY 5 KYAFRAIVVSGSALAGVVPQWGGGNGNGGNS---SGPDSITLSIYQYGSANALALQS 61
Db 376 QVASENLVIISGGTIN--YFTIGGGSATNATNQTISGGKVTSTIYGGNAN-----K 427
QY 62 DARKSETTITQSGYNGADV--GQGADNSTIELTQNGFRNATIOWNAKNSDITVQYQ 119
Db 428 SANENKVITE-GTANVADYGGKSIENNSI-----ANKNSITISGGTLQVTNI---YG 477
QY 120 GNNAAVYNVDQL-----VTRVW-THEMAHANNATAN 149
Db 478 GHSAXDANENSIQISNGGNINNIIVGHAQDHTNLNTIN 515

Search completed: August 2, 2004, 14:54:36
Job time : 31.7 secs

```

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: August 2, 2004, 14:35:42 ; Search time 44.9 Seconds  
(without alignments)  
950.215 Million cell updates/sec

Title: US-09-543-407-14  
Perfect score: 775  
Sequence: 1 MKLLKVAAPAAIVVSGSALA.....HASVMVRQVFGNNATANQY 151

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 1586107 seqs, 282547505 residues

Total number of hits satisfying chosen parameters: 1586107

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : A\_Geneseq\_29Jan04:\*

- 1: Geneseq1980s:\*
- 2: Geneseq1990s:\*
- 3: Geneseq2000s:\*
- 4: Geneseq2001s:\*
- 5: Geneseq2002s:\*
- 6: Geneseq2003as:\*
- 7: Geneseq2003bs:\*
- 8: Geneseq2004s:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	775	100.0	151	3 AAB36347	Aab36347 Agfa::PT3
2	714	92.1	151	3 AAB36352	Aab36352 Agfa::PT3
3	696	89.8	151	3 AAB36346	Aab36346 Agfa::PT3
4	693	89.4	151	2 AAR74625	Aar74625 Agfa sequ
5	693	89.4	151	3 AAB36341	Aab36341 Salmonell
6	692	89.3	151	2 AAB23570	Aab23570 Salmonell
7	659	85.0	151	3 AAB36355	Aab36355 Agfa::PT3
8	612	79.0	151	3 AAB36353	Aab36353 Agfa::PT3
9	611	78.8	151	3 AAB36349	Aab36349 Agfa::PT3
10	609	78.6	151	3 AAB36350	Aab36350 Agfa::PT3
11	603	77.8	151	3 AAB36354	Aab36354 Agfa::PT3
12	602	77.7	151	3 AAB36351	Aab36351 Agfa::PT3
13	578	74.6	151	3 AAB36348	Aab36348 Agfa::PT3
14	528	68.1	151	3 AAB36343	Aab36343 Escherich
15	523	67.5	151	7 ABR82651	Abr82651 E. coli C
16	514	66.3	120	2 AAR62761	Aar62761 Agfa sequ
17	514	66.3	120	2 AAW23569	Aaw23569 Salmonell
18	463	59.7	142	2 AAR52664	Aar52664 Ribronect
19	391	50.5	122	2 AAR52663	Aar52663 FNB curli
20	237	30.6	45	3 AAB36316	Aab36316 Salmonell
21	132	17.0	22	3 AAB36338	Aab36338 Salmonell
22	123	15.9	23	3 AAB36321	Aab36321 Salmonell
23	123	15.9	23	3 AAB36326	Aab36326 Salmonell
24	123	15.9	23	3 AAB36338	Aab36338 Salmonell
25	115	14.8	22	3 AAB36325	Aab36325 Salmonell

26	115	14.8	22	3 AAB36339	Aab36339 Salmonell
27	115	14.8	22	3 AAB36320	Aab36320 Salmonell
28	113	14.6	24	7 ABR82644	Abr82644 E. coli C
29	109	14.1	23	3 AAB36340	Aab36340 Salmonell
30	109	14.1	23	3 AAB36324	Aab36324 Salmonell
31	109	14.1	23	3 AAB36319	Aab36319 Salmonell
32	104.5	13.5	151	3 AAB36344	Aab36344 Escherich
33	98.5	12.7	151	3 AAB36342	Aab36342 Salmonell
34	98	12.6	26	7 ABR82645	Abr82645 E. coli C
35	95	12.3	186	6 ABU21488	Abu21488 Protein e
36	93	12.0	19	3 AAB36323	Aab36323 Salmonell
37	93	12.0	19	3 AAB36336	Aab36336 Salmonell
38	93	12.0	19	3 AAB36328	Aab36328 Salmonell
39	92	11.9	23	3 AAB36331	Aab36331 Escherich
40	92	11.9	673	5 AAU44403	Aau44403 Human tru
41	92	11.9	673	5 AAU79538	Aau79538 Truncated
42	92	11.9	949	3~ AAU44404	Aau44404 Human tru
43	92	11.9	949	5 AAU79539	Aau79539 Truncated
44	92	11.9	1327	3 AAU44402	Aau44402 Human tan
45	92	11.9	1327	3 AAB27212	Aab27212 Human tan

ALIGNMENTS

RESULT 1  
AAB36347  
ID AAB36347 standard; protein; 151 AA.  
XX  
AC AAB36347;  
XX  
DT 26-FEB-2001 (first entry)  
XX  
DE Agfa::PT3#2 amino acid sequence SEQ ID NO:14.  
XX  
KW Salmonella; agfa; chromosomal gene replacement; fimbrin; epitope;  
KW vaccine; immune response; immunogen.  
XX  
OS Salmonella enteritidis.  
OS Escherichia coli.  
OS Synthetic.  
XX  
PN WO200060102-A2.  
XX  
PD 12-OCT-2000.  
XX  
PF 05-APR-2000; 2000WO-CA000356.  
XX  
PR 05-APR-1999; 99US-0127888P.  
(UYVI-) UNIV VICTORIA.  
White AP, Doran JL, Collison SK, Kay WW;  
WPI: 2000-672631/65.  
N-PSDB; AAC64623.

Recombinant agfa gene having a segment replaced by a foreign DNA sequence which encodes foreign epitope or antigen, expresses recombinant Agfa protein useful for eliciting immune response in animal.

Disclosure; Page 136; 139pp; English.

The present invention describes a recombinant agfa gene (I) where a segment of the gene has been replaced by a segment of a foreign DNA sequence which encodes a foreign epitope or antigen. Also described are: (1) use of thin aggregative fimbriae (SFF1/TAF) nucleation depended assembly system of strains of Salmonella, Escherichia coli and Enterobacteriaceae for the production of fimbriae comprising recombinant Agfa, CsgA and Agfa-homologue fimbryn subunits, respectively; (2) directing recombination of a recombinant gene into the chromosome of the homologous species; (3) directing recombination of a recombinant gene back into the chromosome of the homologous species, replacing the native

CC copy of that gene; and (4) eliciting an immune response in an animal,  
 CC comprising separating an amino acid polymer comprising a recombinant Agfa  
 CC protein containing a replacement segment or segments of foreign amino  
 CC acid sequence or sequences grown on a *Salmonella*, *E. coli* or  
 CC Enterobacteriaceae host cell, from the host cell and introducing the  
 CC polymer into the animal in conjunction with a carrier or diluent. (1) is  
 CC useful for the expression of recombinant Agfa protein which is useful for  
 CC eliciting an immune response in an animal. In a fibrial presentation  
 CC system the heterologous antigens are presented in high numbers (up to  
 CC 500,000 copies/cell), the hybrid fimbriae protein possesses both the  
 CC immunogenicity and adhesion properties relevant for an efficient live  
 CC vaccine, the carrier fibrial subunit proteins are usually strong  
 CC immunogens, which may be important for directing an immune response  
 CC against the inserted epitope, and hybrid fimbriae are easy and  
 CC inexpensive to purify in large amount. The present sequence is given in  
 CC the exemplification of the present invention

XX Sequence 151 AA;

Query Match 100.0%; Score 775; DB 3; Length 151;  
 Best Local Similarity 100.0%; Pred. No. 5.4e-67;  
 Matches 151; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MKLLKVAFAAIVVSGSALAGVVPQWGGGNGHNGGNSGPDSTLSIYQYGSANAALALQ 60  
 |||||  
 DB 1 MKLLKVAFAAIVVSGSALAGVVPQWGGGNGHNGGNSGPDSTLSIYQYGSANAALALQ 60  
 |||||  
 QY 61 SDARKSETTITQSGYNGADVGQADNSTIELTQNGFRNNATIDQWNAKNSDITVGYDQ 120  
 |||||  
 DB 61 SDARKSETTITQSGYNGADVGQADNSTIELTQNGFRNNATIDQWNAKNSDITVGYDQ 120  
 |||||  
 QY 121 LVTRVVTHEMAHASVMVRQVFGNNATANOY 151  
 |||||  
 DB 121 LVTRVVTHEMAHASVMVRQVFGNNATANOY 151  
 |||||

# RESULT 2

AAB36352 ID AAB36352 standard; protein; 151 AA.

XX AAB36352;

XX 26-FEB-2001 (first entry)

XX Agfa::PT3#7 amino acid sequence SEQ ID NO:24.

XX *Salmonella*; agfa; chromosomal gene replacement; fimbria; epitope;  
 XX vaccine; immune response; immunogen.

XX *Salmonella enteritidis*.

XX *Escherichia coli*.

XX Synthetic.

XX WO2000060102-A2.

XX 12-OCT-2000.

XX 05-APR-2000; 2000WO-CA000356.

XX 05-APR-1999; 99US-0127888P.

XX (UYVI-) UNIV VICTORIA.

XX White AP, Doran JL, Collison SK, Kay WW;

XX WPI; 2000-672631/65.

XX N-PSDB; AAC64628.

XX Recombinant agfa gene having a segment replaced by a foreign DNA sequence  
 XX which encodes foreign epitope or antigen, expresses recombinant Agfa  
 XX protein useful for eliciting immune response in animal.

XX Disclosure; Page 138; 139pp; English.

XX The present invention describes a recombinant agfa gene (1) where a  
 CC segment of the gene has been replaced by a segment of a foreign DNA  
 CC sequence which encodes a foreign epitope or antigen. Also described are:  
 CC (1) use of thin aggregative fimbriae (SEF17/TAF) nucleation depended  
 CC assembly system of strains of *Salmonella*, *Escherichia coli* and  
 CC Enterobacteriaceae for the production of fimbriae comprising recombinant  
 CC Agfa, CsgA and Agfa-homologue fimbria subunits, respectively; (2)  
 CC directing recombination of a recombinant gene into the chromosome of the  
 CC homologous species; (3) directing recombination of a recombinant gene  
 CC back into the chromosome of the homologous species, replacing the native  
 CC copy of that gene; and (4) eliciting an immune response in an animal,  
 CC comprising separating an amino acid polymer comprising a recombinant Agfa  
 CC protein containing a replacement segment or segments of foreign amino  
 CC acid sequence or sequences grown on a *Salmonella*, *E. coli* or  
 CC Enterobacteriaceae host cell, from the host cell and introducing the  
 CC polymer into the animal in conjunction with a carrier or diluent. (1) is  
 CC useful for the expression of recombinant Agfa protein which is useful for  
 CC eliciting an immune response in an animal. In a fibrial presentation  
 CC system the heterologous antigens are presented in high numbers (up to  
 CC 500,000 copies/cell), the hybrid fimbriae protein possesses both the  
 CC immunogenicity and adhesion properties relevant for an efficient live  
 CC vaccine, the carrier fibrial subunit proteins are usually strong  
 CC immunogens, which may be important for directing an immune response  
 CC against the inserted epitope, and hybrid fimbriae are easy and  
 CC inexpensive to purify in large amount. The present sequence is given in  
 CC the exemplification of the present invention

XX Sequence 151 AA;

Query Match 92.1%; Score 714; DB 3; Length 151;  
 Best Local Similarity 91.1%; Pred. No. 4.3e-61;  
 Matches 144; Conservative 0; Mismatches 0; Indels 14; Gaps 2;

QY 1 MKLLKVAFAAIVVSGSALAGVVPQWGGGNGHNGGNSGPDSTLSIYQYGSANAALALQ 60  
 |||||  
 DB 1 MKLLKVAFAAIVVSGSALAGVVPQWGGGNGHNGGNSGPDSTLSIYQYGSANAALALQ 60  
 |||||  
 QY 61 SDARKSETTITQSGYNGADVGQADNSTIELTQNGFRNNATIDQWNAKNSDITVGYDQ 120  
 |||||  
 DB 61 SDARKSETTITQSGYNGADVGQADNSTIELTQNGFRNNATIDQWNAKNSDITVGYDQ 120  
 |||||  
 QY 121 LVTRVVTHEMAHASVMVRQVFGNNATANOY 151  
 |||||  
 DB 114 LVTRVVTHEMAHASVMVRQVFGNNATANOY 151  
 |||||

# RESULT 3

AAB36346 ID AAB36346 standard; protein; 151 AA.

XX AAB36346;

XX 26-FEB-2001 (first entry)

XX Agfa::PT3#1 amino acid sequence SEQ ID NO:12.

XX *Salmonella*; agfa; chromosomal gene replacement; fimbria; epitope;  
 XX vaccine; immune response; immunogen.

XX *Salmonella enteritidis*.

XX *Escherichia coli*.

XX Synthetic.

XX WO2000060102-A2.

XX 12-OCT-2000.

XX 05-APR-2000; 2000WO-CA000356.

XX 05-APR-1999; 99US-0127888P.

XX (UYVI-) UNIV VICTORIA.

XX White AP, Doran JL, Collison SK, Kay WW;  
 XX WPI; 2000-672631/65.  
 DR N-PSDB; AAC64622.  
 XX Recombinant agfa gene having a segment replaced by a foreign DNA sequence  
 PT which encodes foreign epitope or antigen, expresses recombinant Agfa  
 PT protein useful for eliciting immune response in animal.  
 XX Disclosure; Page 135; 139pp; English.  
 XX The present invention describes a recombinant agfa gene (1) where a  
 CC segment of the gene has been replaced by a segment of a foreign DNA  
 CC segment which encodes a foreign epitope or antigen. Also described are:  
 CC (1) use of thin aggregative fimbriae (SEPI7/TAf) nucleation depended  
 CC assembly system of strains of Salmonella, Escherichia coli and  
 CC Enterobacteriaceae for the production of fimbriae comprising recombinant  
 CC Agfa, CsgA and Agfa-homologue fimbria subunits, respectively; (2)  
 CC directing recombination of a recombinant gene into the chromosome of the  
 CC homologous species; (3) directing recombination of a recombinant gene  
 CC back into the chromosome of the homologous species, replacing the native  
 CC copy of that gene; and (4) eliciting an immune response in an animal,  
 CC comprising separating an amino acid polymer comprising a recombinant Agfa  
 CC protein containing a replacement segment or segments of foreign amino  
 CC acid sequence or sequences grown on a Salmonella, E. coli or  
 CC Enterobacteriaceae host cell, from the host cell and introducing the  
 CC polymer into the animal in conjunction with a carrier or diluent. (1) is  
 CC useful for the expression of recombinant Agfa protein which is useful for  
 CC eliciting an immune response in an animal. In a fimbrial presentation  
 CC system the heterologous antigens are presented in high numbers (up to  
 CC 500,000 copies/cell), the hybrid fimbria protein possesses both the  
 CC immunogenicity and adhesion properties relevant for an efficient live  
 CC vaccine, the carrier fimbrial subunit proteins are usually strong  
 CC immunogens, which may be important for directing an immune response  
 CC against the inserted epitope, and hybrid fimbriae are easy and  
 CC inexpensive to purify in large amount. The present sequence is given in  
 CC the exemplification of the present invention  
 XX Sequence 151 AA;  
 SQ

Query Match 89.8%; Score 696; DB 3; Length 151;  
 Best Local Similarity 87.6%; Pred. No. 2.4e-59;  
 Matches 141; Conservative 0; Mismatches 0; Indels 20; Gaps 2;

QY 1 MKLLKVAAPAAIVVSGSALAGVVPQWGGGNGHNGGNSGDPSTLSIYQYGSANAALALQ 60  
 Db 1 MKLLKVAAPAAIVVSGSALAGVVPQWGGGNGHNGGNSGDPSTLSIYQYGSANAALALQ 60  
 QY 61 SDARKSETTITQSGYNGADVGGQADNSTIELTQNGFRNNATIDQWNAKNSDITVGGQ 117  
 Db 61 SDARKSETTITQSGYNGADVGGQADNSTIELTQNGFRNNATIDQWNAKNSDITVGGQ 120  
 QY 118 -----YDQVTRVTHEMAHASVVMVQVGFNNATANQY 151  
 Db 121 NNAALVNYDQVTRVTHEMAHA-----NNATANQY 151

RESULT 4  
 AAR74625  
 ID AAR74625 standard; protein; 151 AA.  
 XX AAR74625;  
 AC  
 XX 25-MAR-2003 (revised)  
 DT 26-JUN-1995 (first entry)  
 XX Agfa sequence.  
 DE  
 XX Salmonella; Agfa; vaccine.  
 KW  
 XX Salmonella.  
 OS

PN WO9425598-A2.  
 XX 10-NOV-1994.  
 XX 26-APR-1994; 94WO-IB000207.  
 XX 26-APR-1993; 93US-00054452.  
 XX (UYVI-) UNIV VICTORIA INNOVATION & DEV CORP.  
 PA (KING/) KING J.  
 XX Kay WW, Collinson SK, Clouthier SC, Doran JL;  
 DR WPI; 1994-358275/44.  
 DR N-PSDB; AAQ87467.  
 XX Eliciting an immune response to Salmonella - using attenuated Salmonella  
 PT strains, vector constructs, or compans. contg. fimbrial type proteins.  
 PT Disclosure; Fig 7B; 95pp; English.  
 PS The Salmonella Agfa protein and DNA are used in vaccine and genetic  
 CC immunization compositions, respectively, to elicit an immune response to  
 CC Salmonella in animals (e.g. food producing animals) and humans. (Updated  
 CC on 25-MAR-2003 to correct PN field.)  
 XX Sequence 151 AA;  
 SQ

Query Match 89.4%; Score 693; DB 2; Length 151;  
 Best Local Similarity 90.7%; Pred. No. 4.7e-59;  
 Matches 137; Conservative 3; Mismatches 11; Indels 0; Gaps 0;

QY 1 MKLLKVAAPAAIVVSGSALAGVVPQWGGGNGHNGGNSGDPSTLSIYQYGSANAALALQ 60  
 Db 1 MKLLKVAAPAAIVVSGSALAGVVPQWGGGNGHNGGNSGDPSTLSIYQYGSANAALALQ 60  
 QY 61 SDARKSETTITQSGYNGADVGGQADNSTIELTQNGFRNNATIDQWNAKNSDITVGGQ 120  
 Db 61 SDARKSETTITQSGYNGADVGGQADNSTIELTQNGFRNNATIDQWNAKNSDITVGGQ 120  
 QY 121 LVTRVTHEMAHASVVMVQVGFNNATANQY 151  
 Db 121 NNAALVNYDQVTRVTHEMAHA-----NNATANQY 151

RESULT 5  
 AAB36341  
 ID AAB36341 standard; protein; 151 AA.  
 XX AAB36341;  
 AC  
 XX 26-FEB-2001 (first entry)  
 DT  
 XX Salmonella enteritidis Agfa amino acid sequence SEQ ID NO:5.  
 DE  
 XX Salmonella; agfa; chromosomal gene replacement; fimbrin; epitope;  
 KW vaccine; immune response; immunogen.  
 XX Salmonella enteritidis.  
 OS  
 XX WO2000060102-A2.  
 PN  
 XX 12-OCT-2000.  
 PD  
 XX 05-APR-2000; 2000WO-CA000356.  
 PF  
 XX 05-APR-1999; 99US-0127888P.  
 PR  
 XX (UYVI-) UNIV VICTORIA.  
 PA  
 XX White AP, Doran JL, Collison SK, Kay WW;  
 PI WPI; 2000-672631/65.  
 DR

DR N-PSDB; AAC64617.  
 XX Recombinant agfa gene having a segment replaced by a foreign DNA sequence  
 PT which encodes foreign epitope or antigen, expresses recombinant Agfa  
 PT protein useful for eliciting immune response in animal.  
 XX  
 XX Disclosure; Page 135; 139pp; English.  
 XX  
 CC The present invention describes a recombinant agfa gene (I) where a  
 CC segment of the gene has been replaced by a segment of a foreign DNA  
 CC sequence which encodes a foreign epitope or antigen. Also described are:  
 CC (1) use of thin aggregative fimbriae (SEF17/TAF) nucleation depended  
 CC assembly system of strains of *Salmonella*, *Escherichia coli* and  
 CC *Enterobacteriaceae* for the production of fimbriae comprising recombinant  
 CC Agfa, CsgA and Agfa-homologue fimbria subunits, respectively; (2)  
 CC directing recombination of a recombinant gene into the chromosome of the  
 CC homologous species; (3) directing recombination of a recombinant gene  
 CC back into the chromosome of the homologous species, replacing the native  
 CC copy of that gene; and (4) eliciting an immune response in an animal,  
 CC comprising separating an amino acid polymer comprising a recombinant Agfa  
 CC protein containing a replacement segment or segments of foreign amino  
 CC acid sequence or sequences grown on a *Salmonella*, *E. coli* or  
 CC *Enterobacteriaceae* host cell, from the host cell and introducing the  
 CC polymer into the animal in conjunction with a carrier or diluent. (I) is  
 CC useful for the expression of recombinant Agfa protein which is useful for  
 CC eliciting an immune response in an animal. In a fimbrial presentation  
 CC system the heterologous antigens are presented in high numbers (up to  
 CC 500,000 copies/cell), the hybrid fimbria protein possesses both the  
 CC immunogenicity and adhesion properties relevant for an efficient live  
 CC vaccine, the carrier fimbrial subunit proteins are usually strong  
 CC immunogens, which may be important for directing an immune response  
 CC against the inserted epitope, and hybrid fimbriae are easy and  
 CC inexpensive to purify in large amount. The present sequence is given in  
 CC the exemplification of the present invention  
 XX  
 XX Sequence 151 AA;

Query Match 89.4%; Score 693; DB 3; Length 151;  
 Best Local Similarity 90.7%; Pred. No. 4.7e-59;  
 Matches 137; Conservative 3; Mismatches 11; Indels 0; Gaps 0;  
 QY 1 MKLLKVAFAAIVVSGSALAGVVPQWGGGNGHNGGNSGPDSTLSIYQYGSANAALALQ 60  
 DB 1 MKLLKVAFAAIVVSGSALAGVVPQWGGGNGHNGGNSGPDSTLSIYQYGSANAALALQ 60  
 QY 61 SDARKSETTITQSGYNGADVGGQADNSTIELTQNGFRNNATIDQWNAKNSDITVGGYDQ 120  
 DB 61 SDARKSETTITQSGYNGADVGGQADNSTIELTQNGFRNNATIDQWNAKNSDITVGGYDQ 120  
 QY 121 LVTRVVTHEMAHASVMVROVQVFGNNATANQY 151  
 DB 121 NNAALVNQTASDSSVMVROVQVFGNNATANQY 151  
 RESULT 6  
 AAW23570  
 ID AAW23570 standard; protein; 151 AA.  
 AC AAW23570;  
 XX  
 XX 25-MAR-2003 (revised)  
 DT 29-SEP-1997 (first entry)  
 XX  
 XX *Salmonella enteritidis* 27655-3b agfa.  
 DE  
 XX Enteropathogenic bacteria; enterobacteria; *S. enteritidis*; antibody.  
 XX *Salmonella enteritidis*.  
 OS  
 XX Key Location/Qualifiers  
 FH Misc-difference 123  
 FT /note= "Encoded by GCC"  
 FT  
 XX

PN US5635617-A.  
 XX  
 XX 03-JUN-1997.  
 XX  
 XX 26-APR-1994; 94US-00233788.  
 XX  
 XX 26-APR-1993; 93US-00054452.  
 XX  
 XX (UVVI-) UNIV VICTORIA INNOVATION & DEV CORP.  
 XX  
 XX Collinson SK, Kay WW, Doran JL;  
 XX WPI; 1997-309886/28.  
 XX N-PSDB; AAT74142.  
 XX  
 CC Isolated *Salmonella* gene agfa - used for diagnosis of *Salmonella* or  
 CC enteropathogenic bacteria of the Enterobacteriaceae family.  
 PT  
 XX Example 2; Fig 7; 85pp; English.  
 XX  
 CC The present sequence represents agfa encoded by the full agfa gene  
 CC derived from *Salmonella enteritidis* 27655-3b. The nucleic acid can be  
 CC used to provide diagnostic assays for *Salmonella* and/or enteropathogenic  
 CC bacteria of the family Enterobacteriaceae. It can also be used to provide  
 CC proteins and antibodies which can be used for assays. The nucleic acid  
 CC sequence can be used to provide probes or primers which can specifically  
 CC hybridise to nucleic acid molecules from greater than 99% of *Salmonella*  
 CC strains that are pathogenic to warm-blooded animals relative to nucleic  
 CC acid molecules from virtually all other microbial organisms. (Updated on  
 CC 25-MAR-2003 to correct PF field.)  
 XX  
 XX Sequence 151 AA;  
 SQ  
 Query Match 89.3%; Score 692; DB 2; Length 151;  
 Best Local Similarity 90.7%; Pred. No. 5.8e-59;  
 Matches 137; Conservative 3; Mismatches 11; Indels 0; Gaps 0;  
 QY 1 MKLLKVAFAAIVVSGSALAGVVPQWGGGNGHNGGNSGPDSTLSIYQYGSANAALALQ 60  
 DB 1 MKLLKVAFAAIVVSGSALAGVVPQWGGGNGHNGGNSGPDSTLSIYQYGSANAALALQ 60  
 QY 61 SDARKSETTITQSGYNGADVGGQADNSTIELTQNGFRNNATIDQWNAKNSDITVGGYDQ 120  
 DB 61 SDARKSETTITQSGYNGADVGGQADNSTIELTQNGFRNNATIDQWNAKNSDITVGGYDQ 120  
 QY 121 LVTRVVTHEMAHASVMVROVQVFGNNATANQY 151  
 DB 121 NNPALVNQTASDSSVMVROVQVFGNNATANQY 151  
 RESULT 7  
 AAB36355  
 ID AAB36355 standard; protein; 151 AA.  
 XX  
 XX AAB36355;  
 XX  
 XX 26-FEB-2001 (first entry)  
 DT  
 XX  
 XX Agfa::PT3#10 amino acid sequence SEQ ID NO:30.  
 DE  
 XX *Salmonella*; agfa; chromosomal gene replacement; fimbria; epitope;  
 KW vaccine; immune response; immunogen.  
 XX  
 XX *Salmonella enteritidis*.  
 OS  
 XX *Escherichia coli*.  
 OS Synthetic.  
 XX  
 XX WO2000060102-A2.  
 PN  
 XX  
 XX 12-OCT-2000.  
 PD  
 XX  
 XX 05-APR-2000; 2000WO-CA000356.  
 XX

```

PR 05-APR-1999; 99US-0127888P.
PA (UYVI-) UNIV VICTORIA.
XX White AP, Doran JL, Collison SK, Kay WW;
XX WPI: 2000-672631/65.
XX N-PSDB; AAC64631.
XX
XX Recombinant agfa gene having a segment replaced by a foreign DNA sequence
XX which encodes foreign epitope or antigen, expresses recombinant Agfa
XX protein useful for eliciting immune response in animal.
XX
XX Disclosure; Page 139; 139pp; English.
XX
XX The present invention describes a recombinant agfa gene (I) where a
XX segment of the gene has been replaced by a segment of a foreign DNA
XX sequence which encodes a foreign epitope or antigen. Also described are:
XX (1) use of thin aggregative fimbriae (SEF17/TAF) nucleation depended
XX assembly system of strains of Salmonella, Escherichia coli and
XX Enterobacteriaceae for the production of fimbriae comprising recombinant
XX Agfa, CsgA and Agfa-homologue fimbrial subunits, respectively; (2)
XX homologous species; (3) directing recombination of a recombinant gene
XX back into the chromosome of the homologous species, replacing the native
XX copy of that gene; and (4) eliciting an immune response in an animal,
XX comprising separating an amino acid polymer comprising a recombinant Agfa
XX protein containing a replacement segment or segments of foreign amino
XX acid sequence or sequences grown on a Salmonella, E. coli or
XX Enterobacteriaceae host cell, from the host cell and introducing the
XX polymer into the animal in conjunction with a carrier or diluent. (I) is
XX useful for the expression of recombinant Agfa protein which is useful for
XX eliciting an immune response in an animal. In a fimbrial presentation
XX system the heterologous antigens are presented in high numbers (up to
XX 500,000 copies/cell), the hybrid fimbrial protein possesses both the
XX immunogenicity and adhesion properties relevant for an efficient live
XX vaccine, the carrier fimbrial subunit proteins are usually strong
XX immunogens, which may be important for directing an immune response
XX against the inserted epitope, and hybrid fimbriae are easy and
XX inexpensive to purify in large amount. The present sequence is given in
XX the exemplification of the present invention
XX
XX Sequence 151 AA;
XX
XX Query Match 85.0%; Score 659; DB 3; Length 151;
XX Best Local Similarity 81.9%; Pred. No. 9.1e-56;
XX Matches 136; Conservative 0; Mismatches 0; Indels 30; Gaps 2;
XX
XX QY 1 MKLLKVAFAAIVVSGSALAGVVPQWGGGNGHNGGNSGDPSTLSIYQYGSANAALALQ 60
XX DB 1 MKLLKVAFAAIVVSGSALAGVVPQWGGGNGHNGGNSGDPSTLSIYQYGSANAALALQ 60
XX
XX QY 61 SDARKSETTITQSGYNGADVGQADNSTIELTQNGFRNNATIDQWNAKNSDITVGYDQ 120
XX DB 61 SDARKSETTITQSGYNGADVGQADNSTIELTQNGFRNNATIDQWNAKNSDITVGYDQ 120
XX
XX QY 121 LVTRVVTHEMAHAGVQGVGGNHNATANQY 151
XX DB 106 LVTRVVTHEMAHAGVQGVGGNHNATANQY 151
XX
XX RESULT 8
XX AAB36353
XX ID AAB36353 standard; protein; 151 AA.
XX AC AAB36353;
XX XX
XX 26-FEB-2001 (first entry)
XX DE Agfa::PT3#8 amino acid sequence SEQ ID NO:26.
XX XX
XX Salmonella; agfa; chromosomal gene replacement; fimbrial; epitope;
XX vaccine; immune response; immunogen.

```

```

XX Salmonella enteritidis.
XX Escherichia coli.
XX Synthetic.
XX WO2000060102-A2.
XX 12-OCT-2000.
XX
XX 05-APR-2000; 2000WO-CA000356.
XX
XX 05-APR-1999; 99US-0127888P.
XX (UYVI-) UNIV VICTORIA.
XX
XX White AP, Doran JL, Collison SK, Kay WW;
XX WPI: 2000-672631/65.
XX N-PSDB; AAC64629.
XX
XX Recombinant agfa gene having a segment replaced by a foreign DNA sequence
XX which encodes foreign epitope or antigen, expresses recombinant Agfa
XX protein useful for eliciting immune response in animal.
XX
XX Disclosure; Page 138; 139pp; English.
XX
XX The present invention describes a recombinant agfa gene (I) where a
XX segment of the gene has been replaced by a segment of a foreign DNA
XX sequence which encodes a foreign epitope or antigen. Also described are:
XX (1) use of thin aggregative fimbriae (SEF17/TAF) nucleation depended
XX assembly system of strains of Salmonella, Escherichia coli and
XX Enterobacteriaceae for the production of fimbriae comprising recombinant
XX Agfa, CsgA and Agfa-homologue fimbrial subunits, respectively; (2)
XX homologous species; (3) directing recombination of a recombinant gene
XX back into the chromosome of the homologous species, replacing the native
XX copy of that gene; and (4) eliciting an immune response in an animal,
XX comprising separating an amino acid polymer comprising a recombinant Agfa
XX protein containing a replacement segment or segments of foreign amino
XX acid sequence or sequences grown on a Salmonella, E. coli or
XX Enterobacteriaceae host cell, from the host cell and introducing the
XX polymer into the animal in conjunction with a carrier or diluent. (I) is
XX useful for the expression of recombinant Agfa protein which is useful for
XX eliciting an immune response in an animal. In a fimbrial presentation
XX system the heterologous antigens are presented in high numbers (up to
XX 500,000 copies/cell), the hybrid fimbrial protein possesses both the
XX immunogenicity and adhesion properties relevant for an efficient live
XX vaccine, the carrier fimbrial subunit proteins are usually strong
XX immunogens, which may be important for directing an immune response
XX against the inserted epitope, and hybrid fimbriae are easy and
XX inexpensive to purify in large amount. The present sequence is given in
XX the exemplification of the present invention
XX
XX Sequence 151 AA;
XX
XX Query Match 79.0%; Score 612; DB 3; Length 151;
XX Best Local Similarity 81.5%; Pred. No. 3.2e-51;
XX Matches 123; Conservative 5; Mismatches 23; Indels 0; Gaps 0;
XX
XX QY 1 MKLLKVAFAAIVVSGSALAGVVPQWGGGNGHNGGNSGDPSTLSIYQYGSANAALALQ 60
XX DB 1 MKLLKVAFAAIVVSGSALAGVVPQWGGGNGHNGGNSGDPSTLSIYQYGSANAALYDQ 60
XX
XX QY 61 SDARKSETTITQSGYNGADVGQADNSTIELTQNGFRNNATIDQWNAKNSDITVGYDQ 120
XX DB 61 LVTRVVTHEMAHAGVQGVGGNHNATANQY 151
XX
XX QY 121 LVTRVVTHEMAHAGVQGVGGNHNATANQY 151
XX DB 121 NNAALVNOTASDSSVMVRQVQGVGNNATANQY 151
XX
XX RESULT 9

```

AAB36349  
 ID AAB36349 standard; protein; 151 AA.  
 XX  
 AC AAB36349;  
 XX  
 DT 26-FEB-2001 (first entry)  
 XX  
 XX Agfa::PT3#4 amino acid sequence SEQ ID NO:18.  
 DE  
 XX Salmonella; agfa; chromosomal gene replacement; fimbrin; epitope;  
 KW vaccine; immune response; immunogen.  
 XX  
 XX Salmonella enteritidis.  
 OS Escherichia coli.  
 OS Synthetic.  
 OS  
 XX WO200060102-A2.  
 PN  
 XX 12-OCT-2000.  
 PD  
 XX 05-APR-2000; 2000WO-CA000356.  
 PF  
 XX 05-APR-1999; 99US-0127888P.  
 PR  
 XX (UYVI-) UNIV VICTORIA.  
 PA  
 XX White AP, Doran JL, Collison SK, Kay WW;  
 PI WPI; 2000-672631/65.  
 DR N-PSDB; AAC64625.  
 DR  
 XX Recombinant agfa gene having a segment replaced by a foreign DNA sequence  
 PT which encodes foreign epitope or antigen, expresses recombinant Agfa  
 PT protein useful for eliciting immune response in animal.  
 PT  
 XX Disclosure; Page 136; 139pp; English.  
 PS  
 XX The present invention describes a recombinant agfa gene (I) where a  
 CC segment of the gene has been replaced by a segment of a foreign DNA  
 CC sequence which encodes a foreign epitope or antigen. Also described are:  
 CC (1) use of thin aggregative fimbriae (SEF17/TAFF) nucleation depended  
 CC assembly system of strains of Salmonella, Escherichia coli and  
 CC Enterobacteriaceae for the production of fimbriae comprising recombinant  
 CC Agfa, CSga and Agfa-homologue fimbrin subunits, respectively; (2)  
 CC directing recombination of a recombinant gene into the chromosome of the  
 CC homologous species; (3) directing recombination of a recombinant gene  
 CC back into the chromosome of the homologous species, replacing the native  
 CC copy of that gene; and (4) eliciting an immune response in an animal,  
 CC comprising separating an amino acid polymer comprising a recombinant Agfa  
 CC protein containing a replacement segment or segments of foreign amino  
 CC acid sequence or sequences grown on a Salmonella, E. coli or  
 CC Enterobacteriaceae host cell, from the host cell and introducing the  
 CC polymer into the animal in conjunction with a carrier or diluent. (I) is  
 CC useful for the expression of recombinant Agfa protein which is useful for  
 CC eliciting an immune response in an animal. In a fimbrial presentation  
 CC system the heterologous antigens are presented in high numbers (up to  
 CC 500,000 copies/cell), the hybrid fimbrin protein possesses both the  
 CC immunogenicity and adhesion properties relevant for an efficient live  
 CC vaccine, the carrier fimbrial subunit proteins are usually strong  
 CC immunogens, which may be important for directing an immune response  
 CC against the inserted epitope, and hybrid fimbriae are easy and  
 CC inexpensive to purify in large amount. The present sequence is given in  
 CC the exemplification of the present invention  
 XX  
 XX Sequence 151 AA;  
 SQ  
 Query Match 78.8%; Score 611; DB 3; Length 151;  
 Best Local Similarity 81.5%; Pred. No. 4e-51;  
 Matches 123; Conservative 6; Mismatches 22; Indels 0; Gaps 0;  
 1 MKLLKVAAPAAIYVSGSALAGVVPQWGGGNGHNGGNSGSPDSTLSIYGYGSANAALAQ 60  
 1 MKLLKVAAPAAIYVSGSALAGVVPQWGGGNGHNGGNSGSPDSTLSIYGYGSANAALAQ 60

61 SDARKSETTITQSGYNGADVGGADNSTIELTQNGFRNNAITDQWNAKNSDITVQYDQ 120  
 61 SDARKSETTITQSGYNGADVGGADNSTIELTQNGFRNNAITDQWNAKNSDITVQYGG 120  
 121 LVTRVVTHEMAHASVMVRQVGFNNATANQY 151  
 121 NNAALVNQATASDSSVMVRQVGFNNATANQY 151  
 RESULT 10  
 AAB36350  
 ID AAB36350 standard; protein; 151 AA.  
 XX  
 AC AAB36350;  
 XX  
 DT 26-FEB-2001 (first entry)  
 XX  
 XX Agfa::PT3#5 amino acid sequence SEQ ID NO:20.  
 DE  
 XX Salmonella; agfa; chromosomal gene replacement; fimbrin; epitope;  
 KW vaccine; immune response; immunogen.  
 XX  
 XX Salmonella enteritidis.  
 OS Escherichia coli.  
 OS Synthetic.  
 OS  
 XX WO200060102-A2.  
 PN  
 XX 12-OCT-2000.  
 PD  
 XX 05-APR-2000; 2000WO-CA000356.  
 PF  
 XX 05-APR-1999; 99US-0127888P.  
 PR  
 XX (UYVI-) UNIV VICTORIA.  
 PA  
 XX White AP, Doran JL, Collison SK, Kay WW;  
 PI WPI; 2000-672631/65.  
 DR N-PSDB; AAC64626.  
 DR  
 XX Recombinant agfa gene having a segment replaced by a foreign DNA sequence  
 PT which encodes foreign epitope or antigen, expresses recombinant Agfa  
 PT protein useful for eliciting immune response in animal.  
 PT  
 XX Disclosure; Page 137; 139pp; English.  
 PS  
 XX The present invention describes a recombinant agfa gene (I) where a  
 CC segment of the gene has been replaced by a segment of a foreign DNA  
 CC sequence which encodes a foreign epitope or antigen. Also described are:  
 CC (1) use of thin aggregative fimbriae (SEF17/TAFF) nucleation depended  
 CC assembly system of strains of Salmonella, Escherichia coli and  
 CC Enterobacteriaceae for the production of fimbriae comprising recombinant  
 CC Agfa, CSga and Agfa-homologue fimbrin subunits, respectively; (2)  
 CC directing recombination of a recombinant gene into the chromosome of the  
 CC homologous species; (3) directing recombination of a recombinant gene  
 CC back into the chromosome of the homologous species, replacing the native  
 CC copy of that gene; and (4) eliciting an immune response in an animal,  
 CC comprising separating an amino acid polymer comprising a recombinant Agfa  
 CC protein containing a replacement segment or segments of foreign amino  
 CC acid sequence or sequences grown on a Salmonella, E. coli or  
 CC Enterobacteriaceae host cell, from the host cell and introducing the  
 CC polymer into the animal in conjunction with a carrier or diluent. (I) is  
 CC useful for the expression of recombinant Agfa protein which is useful for  
 CC eliciting an immune response in an animal. In a fimbrial presentation  
 CC system the heterologous antigens are presented in high numbers (up to  
 CC 500,000 copies/cell), the hybrid fimbrin protein possesses both the  
 CC immunogenicity and adhesion properties relevant for an efficient live  
 CC vaccine, the carrier fimbrial subunit proteins are usually strong  
 CC immunogens, which may be important for directing an immune response  
 CC against the inserted epitope, and hybrid fimbriae are easy and  
 CC inexpensive to purify in large amount. The present sequence is given in  
 CC the exemplification of the present invention  
 XX  
 XX Sequence 151 AA;  
 SQ



CC the exemplification of the present invention  
 XX Sequence 151 AA;  
 SQ

Query Match 78.6%; Score 609; DB 3; Length 151;  
 Best Local Similarity 81.5%; Pred. No. 6.3e-51;  
 Matches 123; Conservative 6; Mismatches 22; Indels 0; Gaps 0;

QY 1 MLLKVAFAAIVVSGSALAGVVPQGGGNGHNGSGPDSLTSLIYQYGSANAALALQ 60  
 DB 1 MLLKVAFAAIVVSGSALAGVVPQGGGNGHNGSGPDSLTSLIYQYGSANAALALQ 60  
 QY 61 SDARKSETTITQSGYNGADVQCGADNSTIELTQNGFRNNATIDOWNAKNSDITVGYDQ 120  
 DB 61 SDARKYDQLVTRVVTHEMAHAGGADNSTIELTQNGFRNNATIDOWNAKNSDITVGYGG 120  
 QY 121 LVTRVVTHEMAHASVMVRQVGFNNATANQY 151  
 DB 121 NNAALVNOTASDSSVMVRQVGFNNATANQY 151

RESULT 11  
 AAB36354  
 ID AAB36354 standard; protein; 151 AA.  
 XX AC AAB36354;  
 XX DT 26-FEB-2001 (first entry)  
 XX DE Agfa::PT3#9 amino acid sequence SEQ ID NO:28.  
 XX KW Salmonella; agfa; chromosomal gene replacement; fimbrin; epitope;  
 XX KW vaccine; immune response; immunogen.  
 XX OS Salmonella enteritidis.  
 XX OS Escherichia coli.  
 XX OS Synthetic.  
 XX PN WO200060102-A2.  
 XX PD 12-OCT-2000.  
 XX PF 05-APR-2000; 2000WO-CA000356.  
 XX PR 05-APR-1999; 99US-0127888P.  
 XX PA (UYVI-) UNIV VICTORIA.  
 XX PI White AP, Doran JL, Collison SK, Kay WW;  
 XX DR WPI; 2000-672631/65.  
 XX DR N-PSDB; AAC64630.

Recombinant agfa gene having a segment replaced by a foreign DNA sequence which encodes foreign epitope or antigen, expresses recombinant Agfa protein useful for eliciting immune response in animal.  
 PS Disclosure; Page 138; 139pp; English.

The present invention describes a recombinant agfa gene (I) where a segment of the gene has been replaced by a segment of a foreign DNA sequence which encodes a foreign epitope or antigen. Also described are:  
 CC (1) use of thin aggregative fimbriae (SEPI7/TAF) nucleation depended assembly system of strains of Salmonella, Escherichia coli and Enterobacteriaceae for the production of fimbriae comprising recombinant Agfa, CsgA and Agfa-homologue fimbrial subunits, respectively; (2) directing recombination of a recombinant gene into the chromosome of the homologous species; (3) directing recombination of a recombinant gene back into the chromosome of the homologous species, replacing the native copy of that gene; and (4) eliciting an immune response in an animal, comprising separating an amino acid polymer comprising a recombinant Agfa protein containing a replacement segment or segments of foreign amino acid sequence or sequences grown on a Salmonella, E. coli or

CC Enterobacteriaceae host cell, from the host cell and introducing the polymer into the animal in conjunction with a carrier or diluent. (I) is useful for the expression of recombinant agfa protein which is useful for eliciting an immune response in an animal. In a fimbrial presentation system the heterologous antigens are presented in high numbers (up to 500,000 copies/cell), the hybrid fimbrial protein possesses both the immunogenicity and adhesion properties relevant for an efficient live vaccine, the carrier fimbrial subunit proteins are usually strong immunogens, which may be important for directing an immune response against the inserted epitope, and hybrid fimbriae are easy and inexpensive to purify in large amount. The present sequence is given in the exemplification of the present invention  
 XX Sequence 151 AA;  
 SQ

Query Match 77.8%; Score 603; DB 3; Length 151;  
 Best Local Similarity 80.8%; Pred. No. 2.4e-50;  
 Matches 122; Conservative 4; Mismatches 25; Indels 0; Gaps 0;

QY 1 MLLKVAFAAIVVSGSALAGVVPQGGGNGHNGSGPDSLTSLIYQYGSANAALALQ 60  
 DB 1 MLLKVAFAAIVVSGSALAGVVPQGGGNGHNGSGPDSLTSLIYQYGSANAALALQ 60  
 QY 61 SDARKSETTITQSGYNGADVQCGADNSTIELTQNGFRNNATIDOWNAKNSDITVGYDQ 120  
 DB 61 SDARKSETTITQSGYNGADYDQLVTRVVTHEMAHAFRNNATIDOWNAKNSDITVGYGG 120  
 QY 121 LVTRVVTHEMAHASVMVRQVGFNNATANQY 151  
 DB 121 NNAALVNOTASDSSVMVRQVGFNNATANQY 151

RESULT 12  
 AAB36351  
 ID AAB36351 standard; protein; 151 AA.  
 XX AC AAB36351;  
 XX DT 26-FEB-2001 (first entry)  
 XX DE Agfa::PT3#6 amino acid sequence SEQ ID NO:22.  
 XX KW Salmonella; agfa; chromosomal gene replacement; fimbrin; epitope;  
 XX KW vaccine; immune response; immunogen.  
 XX OS Salmonella enteritidis.  
 XX OS Escherichia coli.  
 XX OS Synthetic.  
 XX PN WO200060102-A2.  
 XX PD 12-OCT-2000.  
 XX PF 05-APR-2000; 2000WO-CA000356.  
 XX PR 05-APR-1999; 99US-0127888P.  
 XX PA (UYVI-) UNIV VICTORIA.  
 XX PI White AP, Doran JL, Collison SK, Kay WW;  
 XX DR WPI; 2000-672631/65.  
 XX DR N-PSDB; AAC64627.

Recombinant agfa gene having a segment replaced by a foreign DNA sequence which encodes foreign epitope or antigen, expresses recombinant Agfa protein useful for eliciting immune response in animal.  
 PS Disclosure; Page 137; 139pp; English.

The present invention describes a recombinant agfa gene (I) where a segment of the gene has been replaced by a segment of a foreign DNA sequence which encodes a foreign epitope or antigen. Also described are:

CC (1) use of thin aggregative fimbriae (SEF17/TAF) nucleation depended  
 CC assembly system of strains of Salmonella, Escherichia coli and  
 CC Enterobacteriaceae for the production of fimbriae comprising recombinant  
 CC AgfA, CsgA and AgfA-homologue fimbrial subunits, respectively; (2)  
 CC directing recombination of a recombinant gene into the chromosome of the  
 CC homologous species; (3) directing recombination of a recombinant gene  
 CC back into the chromosome of the homologous species, replacing the native  
 CC copy of that gene; and (4) eliciting an immune response in an animal,  
 CC comprising separating an amino acid polymer comprising a recombinant AgfA  
 CC protein containing a replacement segment or segments of foreign amino  
 CC acid sequence or sequences grown on a Salmonella, E. coli or  
 CC Enterobacteriaceae host cell, from the host cell and introducing the  
 CC polymer into the animal in conjunction with a carrier or diluent. (1) is  
 CC useful for the expression of recombinant AgfA protein which is useful for  
 CC eliciting an immune response in an animal. In a fimbrial presentation  
 CC system the heterologous antigens are presented in high numbers (up to  
 CC 500,000 copies/cell), the hybrid fimbrial protein possesses both the  
 CC immunogenicity and adhesion properties relevant for an efficient live  
 CC vaccine, the carrier fimbrial subunit proteins are usually strong  
 CC immunogens, which may be important for directing an immune response  
 CC against the inserted epitope, and hybrid fimbriae are easy and  
 CC inexpensive to purify in large amount. The present sequence is given in  
 CC the exemplification of the present invention  
 XX  
 SQ Sequence 151 AA;

Query Match 77.7%; Score 602; DB 3; Length 151;  
 Best Local Similarity 81.5%; Pred. No. 3e-50; Indels 0; Gaps 0;  
 Matches 121; Conservative 5; Mismatches 23;  
 QY 1 M K L L K V A A F A I V V G S A L A G V V P Q G G G N H G G N S G P D S T L S I Y Q Y G S A N A L A L Q 60  
 D b 1 M K L L K V A A F A I V V G S A L A G V V P Q G G G N H G G N S G P D S T L S I Y Q Y G S A N A L A L Q 60  
 QY 61 S D A R K S E T T I T Q S G Y G N G A D V G Q G A D N S T I E L T Q N G F R N N A T I D Q W A K N S D I T V G Q Y D Q 120  
 D b 61 S D A R K S E T T I T Q S G Y G N G A D V G Q G A D N S T I E L T Q N G F R N N A T I D Q W A K N S D I T V G Q Y D Q 120  
 QY 121 L V T R V V T H E M A H A S V M V R Q V G F G N N A T A N Q Y 151  
 D b 121 N N A A L V N Q T A S D S S V M V R Q V G F G N N A T A N Q Y 151

RESULT 13  
 AAB36348  
 ID AAB36348 standard; protein; 151 AA.  
 XX  
 AC AAB36348;  
 XX  
 DT 26-FEB-2001 (first entry)  
 XX  
 DE AgfA::PT3#3 amino acid sequence SEQ ID NO:16.  
 XX  
 KW Salmonella; agfA; chromosomal gene replacement; fimbria; epitope;  
 KW vaccine; immune response; immunogen.  
 XX  
 OS Salmonella enteritidis.  
 OS Escherichia coli.  
 OS Synthetic.  
 XX  
 PN WO2000060102-A2.  
 XX  
 PD 12-OCT-2000.  
 XX  
 PF 05-APR-2000; 2000WO-CR000356.  
 XX  
 PR 05-APR-1999; 99US-0127888P.  
 XX  
 PA (UYVI-) UNIV VICTORIA.  
 XX  
 FI White AP, Doran JL, Collison SK, Kay WH;  
 XX WPI; 2000-672631/65.  
 DR

DR N-PSDB; AAC64624.  
 XX  
 PT Recombinant agfA gene having a segment replaced by a foreign DNA sequence  
 PT which encodes foreign epitope or antigen, expresses recombinant AgfA  
 PT protein useful for eliciting immune response in animal.  
 XX  
 PS Disclosure; Page 136; 139pp; English.

XX The present invention describes a recombinant agfA gene (I) where a  
 CC segment of the gene has been replaced by a segment of a foreign DNA  
 CC sequence which encodes a foreign epitope or antigen. Also described are:  
 CC (1) use of thin aggregative fimbriae (SEF17/TAF) nucleation depended are:  
 CC assembly system of strains of Salmonella, Escherichia coli and  
 CC Enterobacteriaceae for the production of fimbriae comprising recombinant  
 CC AgfA, CsgA and AgfA-homologue fimbrial subunits, respectively; (2)  
 CC directing recombination of a recombinant gene into the chromosome of the  
 CC homologous species; (3) directing recombination of a recombinant gene  
 CC back into the chromosome of the homologous species, replacing the native  
 CC copy of that gene; and (4) eliciting an immune response in an animal,  
 CC comprising separating an amino acid polymer comprising a recombinant AgfA  
 CC protein containing a replacement segment or segments of foreign amino  
 CC acid sequence or sequences grown on a Salmonella, E. coli or  
 CC Enterobacteriaceae host cell, from the host cell and introducing the  
 CC polymer into the animal in conjunction with a carrier or diluent. (1) is  
 CC useful for the expression of recombinant AgfA protein which is useful for  
 CC eliciting an immune response in an animal. In a fimbrial presentation  
 CC system the heterologous antigens are presented in high numbers (up to  
 CC 500,000 copies/cell), the hybrid fimbrial protein possesses both the  
 CC immunogenicity and adhesion properties relevant for an efficient live  
 CC vaccine, the carrier fimbrial subunit proteins are usually strong  
 CC immunogens, which may be important for directing an immune response  
 CC against the inserted epitope, and hybrid fimbriae are easy and  
 CC inexpensive to purify in large amount. The present sequence is given in  
 CC the exemplification of the present invention  
 XX

SQ Sequence 151 AA;

Query Match 74.6%; Score 578; DB 3; Length 151;  
 Best Local Similarity 80.8%; Pred. No. 6.3e-48;  
 Matches 122; Conservative 6; Mismatches 23; Indels 0; Gaps 0;  
 QY 1 M K L L K V A A F A I V V G S A L A G V V P Q G G G N H G G N S G P D S T L S I Y Q Y G S A N A L A L Q 60  
 D b 1 M K L L K V A A F A I V V G S A L A G V V P Q G G G N H G G N S G P D S T L S I Y Q Y G S A N A L A L Q 60  
 QY 61 S D A R K S E T T I T Q S G Y G N G A D V G Q G A D N S T I E L T Q N G F R N N A T I D Q W A K N S D I T V G Q Y D Q 120  
 D b 61 S D A R K S E T T I T Q S G Y G N G A D V G Q G A D N S T I E L T Q N G F R N N A T I D Q W A K N S D I T V G Q Y D Q 120  
 QY 121 L V T R V V T H E M A H A S V M V R Q V G F G N N A T A N Q Y 151  
 D b 121 N N A A L V N Q T A S D S S V M V R Q V G F G N N A T A N Q Y 151

RESULT 14  
 AAB36343  
 ID AAB36343 standard; protein; 151 AA.  
 XX  
 AC AAB36343;  
 XX  
 DT 26-FEB-2001 (first entry)  
 XX  
 DE Escherichia coli CsgA amino acid sequence SEQ ID NO:7.  
 XX  
 KW Salmonella; agfA; chromosomal gene replacement; fimbria; epitope;  
 KW vaccine; immune response; immunogen.  
 XX  
 OS Escherichia coli.  
 OS WO2000060102-A2.  
 XX  
 PN 12-OCT-2000.  
 XX

PF 05-APR-2000; 2000WO-CA000356.  
XX  
PR 05-APR-1999; 99US-0127888P.  
XX  
PA (UUVI-) UNIV VICTORIA.  
XX  
PI White AP, Doran JL, Collison SK, Kay WW;  
XX  
XX WPI; 2000-672631/65.  
DR N-PSDB; AAC64619.  
XX  
XX Recombinant agfa gene having a segment replaced by a foreign DNA sequence  
PT which encodes foreign epitope or antigen, expresses recombinant Agfa  
PT protein useful for eliciting immune response in animal.  
XX  
XX Disclosure; Page 135; 139pp; English.  
XX  
XX The present invention describes a recombinant agfa gene (I) where a  
CC segment of the gene has been replaced by a segment of a foreign DNA  
CC sequence which encodes a foreign epitope or antigen. Also described are:  
CC (1) use of thin aggregative fimbriae (SEF7/TAF) nucleation depended  
CC assembly system of strains of Salmonella, Escherichia coli and  
CC Enterobacteriaceae for the production of fimbriae comprising recombinant  
CC Agfa, CsgA and Agfa-homologue fimbrial subunits, respectively; (2)  
CC directing recombination of a recombinant gene into the chromosome of the  
CC homologous species; (3) directing recombination of a recombinant gene  
CC back into the chromosome of the homologous species, replacing the native  
CC copy of that gene; and (4) eliciting an immune response in an animal,  
CC comprising separating an amino acid polymer comprising a recombinant Agfa  
CC protein containing a replacement segment or segments of foreign amino  
CC acid sequence or sequences grown on a Salmonella, E. coli or  
CC Enterobacteriaceae host cell, from the host cell and introducing the  
CC polymer into the animal in conjunction with a carrier or diluent. (I) is  
CC useful for the expression of recombinant Agfa protein which is useful for  
CC eliciting an immune response in an animal. In a fimbrial presentation  
CC system the heterologous antigens are presented in high numbers (up to  
CC 500,000 copies/cell), the hybrid fimbrial protein possesses both the  
CC immunogenicity and adhesion properties relevant for an efficient live  
CC vaccine, the carrier fimbrial subunit proteins are usually strong  
CC immunogens, which may be important for directing an immune response  
CC against the inserted epitope, and hybrid fimbriae are easy and  
CC inexpensive to purify in large amount. The present sequence is given in  
CC the exemplification of the present invention  
XX  
XX SQ Sequence 151 AA;  
Query Match 68.1%; Score 528; DB 3; Length 151;  
Best Local Similarity 68.9%; Pred. No. 4.3e-43;  
Matches 104; Conservative 20; Mismatches 27; Indels 0; Gaps 0;  
QY 1 MKLLKVAAPAAIVVSGSALAGVVPQWGGGNGHGGNSGPDSTLSIYQGSANALALQ 60  
Db 1 MKLLKVAAPAAIVFSSGALAGVVPQYGGGNGHGGGNSGPNSELNIYQGGNSALALQ 60  
QY 61 SDARKSETTITQSGYNGADVGGADNSTIELTQNGFRNNATIDOWNAKNSDITVGYDQ 120  
Db 61 TDARNSDLTITQHGGENGADVGGSDSSIDLITQRFNGSATLDWNGKNSMTVKQFGG 120  
QY 121 LVTRVVTHEMAHASVMYRQVGFNNATANQY 151  
Db 121 NGGAADVDTASNSSVNVTVQVGFNNATAHQY 151  
RESULT 15  
ID ABR82651  
XX ABR82651 standard; protein; 151 AA.  
XX AC ABR82651;  
XX DT 04-DEC-2003 (first entry)  
XX DE E. coli CsgA subunit 15 kDa protein.  
XX

KW Plasma protein; immune response; antibacterial; vaccine; gene therapy.  
XX Escherichia coli.  
OS WO2003064446-A2.  
XX  
XX 07-AUG-2003.  
XX  
XX 30-JAN-2003; 2003WO-EP000943.  
XX  
XX 31-JAN-2002; 2002GB-00002275.  
XX  
XX (HANS-) HANSA MEDICAL RES AB.  
PA  
XX Bjoerck L, Olsen A, Wikstroem M, Herwald H;  
PI  
XX WPI; 2003-646136/61.  
DR N-PSDB; ACF36153.  
XX  
XX New isolated peptide capable of binding a mammalian plasma protein,  
PT useful in the manufacture of a medicament for the prevention and/or  
PT treatment of a bacterial infection, such as Escherichia coli, Salmonella  
PT or Shigella infections.  
XX  
XX Disclosure; Page 41-42; 42pp; English.  
XX  
XX The invention relates to an isolated peptide capable of binding a  
CC mammalian plasma protein or of generating an immune response in a mammal  
CC selected from sequences shown in ABR82642, ABR82648-49. The peptide or  
CC antibody is useful for treating a bacterial infection in a human or  
CC animal or in the manufacture of a medicament for the prophylactic  
CC treatment of a bacterial infection, such as Escherichia coli, Salmonella  
CC or Shigella infection. The peptide that is immobilized on a solid support  
CC is also useful as a reagent for determining the ability of a plasma  
CC protein to bind to bacteria. The present sequence represents an E. coli  
CC 15 kDa protein  
XX  
XX SQ Sequence 151 AA;  
Query Match 67.5%; Score 523; DB 7; Length 151;  
Best Local Similarity 68.2%; Pred. No. 1.3e-42;  
Matches 103; Conservative 20; Mismatches 28; Indels 0; Gaps 0;  
QY 1 MKLLKVAAPAAIVVSGSALAGVVPQWGGGNGHGGNSGPDSTLSIYQGSANALALQ 60  
Db 1 MKLLKVEAIAAIVFSSGALAGVVPQYGGGNGHGGGNSGPNSELNIYQGGNSALALQ 60  
QY 61 SDARKSETTITQSGYNGADVGGADNSTIELTQNGFRNNATIDOWNAKNSDITVGYDQ 120  
Db 61 TDARNSDLTITQHGGENGADVGGSDSSIDLITQRFNGSATLDWNGKNSMTVKQFGG 120  
QY 121 LVTRVVTHEMAHASVMYRQVGFNNATANQY 151  
Db 121 NGGAADVDTASNSSVNVTVQVGFNNATAHQY 151  
Search completed: August 2, 2004, 14:48:24  
Job time : 44.9 secs



GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: August 2, 2004, 14:40:48 ; Search time 12 Seconds  
(without alignments)  
649.627 Million cell updates/sec

Title: US-09-543-407-14  
Perfect score: 775  
Sequence: 1 MKLLKVAAPAAIVVSGSALA.....HASVWVRQVGFGNATANCY 151

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 389414 seqs, 51625971 residues

Total number of hits satisfying chosen parameters: 389414

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Issued Patents AA:\*  
1: /cgn2\_6/prodata/2/iaa/5A COMB.pep.\*  
2: /cgn2\_6/prodata/2/iaa/5B COMB.pep.\*  
3: /cgn2\_6/prodata/2/iaa/6A COMB.pep.\*  
4: /cgn2\_6/prodata/2/iaa/6B COMB.pep.\*  
5: /cgn2\_6/prodata/2/iaa/PCRUS COMB.pep.\*  
6: /cgn2\_6/prodata/2/iaa/backfiles1.pep.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	692	89.3	151	1	US-08-233-788A-59
2	514	66.3	120	1	US-08-233-788A-57
3	92	11.9	673	3	US-09-196-387-8
4	92	11.9	673	4	US-09-841-835-8
5	92	11.9	949	3	US-09-196-387-10
6	92	11.9	949	4	US-09-841-835-10
7	92	11.9	1327	3	US-09-196-387-2
8	92	11.9	1327	4	US-09-841-835-2
9	92	11.9	1327	4	US-09-372-115A-8
10	89.5	11.5	738	3	US-08-864-038A-3
11	87	11.2	2123	3	US-08-368-685A-10
12	83.5	10.8	339	4	US-09-352-991A-32096
13	83.5	10.8	943	4	US-09-056-556-204
14	83.5	10.8	943	4	US-09-072-596-199
15	83.5	10.8	943	4	US-09-477-135A-131
16	83.5	10.8	943	4	US-09-072-367-204
17	82.5	10.6	2736	4	US-09-352-991A-30227
18	81	10.5	878	4	US-09-540-236-3401
19	80.5	10.4	873	4	US-09-336-447A-13
20	80	10.3	892	4	US-09-336-447A-5
21	79.5	10.3	204	6	US187262-1
22	79.5	10.3	518	3	US-09-043-123-2
23	79.5	10.3	1207	4	US-09-489-039A-11518
24	78.5	10.1	539	4	US-09-719-402A-2
25	78.5	10.1	714	4	US-09-841-786-4
26	78.5	10.1	1912	1	US-08-409-995-4
27	78.5	10.1	1912	3	US-08-685-467-4

28	78.5	10.1	2353	3	US-09-377-155-33	Sequence 33, Appl
29	78.5	10.1	2353	3	US-08-913-942-4	Sequence 4, Appl
30	78.5	10.1	2353	4	US-09-669-974-33	Sequence 33, Appl
31	78.5	10.1	2353	4	US-09-797-862-33	Sequence 33, Appl
32	78.5	10.1	2354	4	US-09-268-347-47	Sequence 47, Appl
33	78.5	10.1	2411	4	US-09-268-347-36	Sequence 36, Appl
34	78.5	10.1	3241	4	US-09-841-786-1	Sequence 1, Appl
35	77.5	10.0	318	4	US-09-107-532A-5818	Sequence 5818, Ap
36	77.5	10.0	3169	4	US-09-453-702B-257	Sequence 257, App
37	77	9.9	415	4	US-09-025-769B-280	Sequence 280, App
38	77	9.9	1690	4	US-09-595-880A-42	Sequence 39, Appl
39	76.5	9.9	238	4	US-09-495-684B-39	Sequence 42, Appl
40	76.5	9.9	956	4	US-09-134-078-63	Sequence 63, Appl
41	76.5	9.9	1413	4	US-09-252-991A-21268	Sequence 21268, A
42	75.5	9.7	645	4	US-09-519-172-41	Sequence 41, Appl
43	75.5	9.7	789	3	US-08-960-780-6	Sequence 6, Appl
44	75.5	9.7	789	3	US-09-073-898-6	Sequence 6, Appl
45	75.5	9.7	789	4	US-09-850-351A-6	Sequence 6, Appl

ALIGNMENTS

RESULT 1  
US-08-233-788A-59  
; Sequence 59, Application US/08233788A  
; Patent No 5635617  
; GENERAL INFORMATION:  
; APPLICANT: Doran, James L.  
; APPLICANT: Kay, William W.  
; APPLICANT: Collinson, Karen S.  
; APPLICANT: Clouthier, Sharon C.  
; TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR DETECTION  
; TITLE OF INVENTION: OF SALMONELLA  
; NUMBER OF SEQUENCES: 61  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Seed and Berry  
; STREET: 6300 Columbia Center, 701 Fifth Avenue  
; CITY: Seattle  
; STATE: Washington  
; COUNTRY: U.S.A.  
; ZIP: 98104-7092  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA: US/08/233,788A  
; APPLICATION NUMBER: US/08/233,788A  
; FILING DATE: 26-APR-1994  
; CLASSIFICATION: 435  
; ATTORNEY/AGENT INFORMATION:  
; NAME: King, Joshua  
; REGISTRATION NUMBER: 35,570  
; REFERENCE/DOCKET NUMBER: 920043.403C2  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (206) 622-4900  
; TELEFAX: (206) 682-6031  
; TELEX: 3723836 SEEDANBERRY  
; INFORMATION FOR SEQ ID NO: 59:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 151 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
US-08-233-788A-59

Query Match 89.3%; Score 692; DB 1; Length 151;  
Best Local Similarity 90.7%; Pred. No. 3.9e-62;  
Matches 137; Conservative 3; Mismatches 11; Indels 0; Gaps 0;

QY 1 MKLLKVAAPAAIVVSGSALAGVVPQWGGNHNGGNSGPDSTLSIYYQSAALALQ 60  
|||||

Db 1 MLLKVAFAAIVVSGSALAGVVPQWGGGNGHNGSSGPDSTLSIYQGSANAALALQ 60  
QY 61 SDARKSETTITQSGYNGADVGQADNSTIETQNGFRNNATIDOWNAKNSDITVGYDQ 120  
Db 61 SDARKSETTITQSGYNGADVGQADNSTIETQNGFRNNATIDOWNAKNSDITVGYGG 120  
QY 121 LVTRVVTHEMAHASVVMVRQVGFNNATANQY 151  
Db 121 NNPALVNTASDSSVMVRQVGFNNATANQY 151

RESULT 2  
US-08-233-788A-57  
; Sequence 57, Application US/08233788A  
; Patent No. 5635617  
; GENERAL INFORMATION:  
; APPLICANT: Doran, James L.  
; APPLICANT: Kay, William W.  
; APPLICANT: Collinson, Karen S.  
; APPLICANT: Clouthier, Sharon C.  
; TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR DETECTION  
; TITLE OF INVENTION: OF SALMONELLA  
; NUMBER OF SEQUENCES: 61  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Seed and Berry  
; STREET: 6300 Columbia Center, 701 Fifth Avenue  
; CITY: Seattle  
; STATE: Washington  
; COUNTRY: U.S.A.  
; ZIP: 98104-7092  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/233,788A  
; FILING DATE: 26-APR-1994  
; CLASSIFICATION: 435  
; ATTORNEY/AGENT INFORMATION:  
; NAME: King, Joshua  
; REGISTRATION NUMBER: 35,570  
; REFERENCE/DOCKET NUMBER: 920043.403C2  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (206) 622-4900  
; TELEFAX: (206) 682-6031  
; TELEX: 3723836 SEEDANBERRY  
; INFORMATION FOR SEQ ID NO: 57:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 120 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
US-08-233-788A-57

Query Match 66.3%; Score 514; DB 1; Length 120;  
Best Local Similarity 100.0%; Pred. No. 2.2e-44;  
Matches 97; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 22 VVPQWGGGNGHNGSSGPDSTLSIYQGSANAALALQSDARKSETTITQSGYNGADV 81  
Db 1 VVPQWGGGNGHNGSSGPDSTLSIYQGSANAALALQSDARKSETTITQSGYNGADV 60  
QY 82 GQADNSTIETQNGFRNNATIDOWNAKNSDITVGY 118  
Db 61 GQADNSTIETQNGFRNNATIDOWNAKNSDITVGY 97

RESULT 3  
US-09-196-387-8  
; Sequence 8, Application US/09196387  
; Patent No. 6277613  
; GENERAL INFORMATION:

; APPLICANT: de Lange, Titia  
; APPLICANT: Smith, Susan  
; TITLE OF INVENTION: A PROTEIN THAT BINDS TO TRF1 AND METHODS  
; TITLE OF INVENTION: OF USE THEREOF  
; NUMBER OF SEQUENCES: 12  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Klauber & Jackson  
; STREET: 411 Hackensack Avenue, 4th Floor  
; CITY: Hackensack  
; STATE: New Jersey  
; COUNTRY: USA  
; ZIP: 07601

; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/09/196,387  
; FILING DATE:

; CLASSIFICATION:  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: 09/095,225  
; FILING DATE: June 10, 1998  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Jackson Esq., David A.  
; REGISTRATION NUMBER: 26,742  
; REFERENCE/DOCKET NUMBER: 600-1-230 CIP1  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 201-487-5800  
; TELEFAX: 201-343-1684  
; TELEX: 133521

; INFORMATION FOR SEQ ID NO: 8:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 673 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
US-09-196-387-8

Query Match 11.9%; Score 92; DB 3; Length 673;  
Best Local Similarity 30.4%; Pred. No. 0.54; Indels 8; Gaps 4;  
Matches 35; Conservative 15; Mismatches 57;  
QY 6 VAAFAAI-VWGSALAGVVPQWGGGNGHNGSSGPDSTLSIYQGSANAALALQSDAR 64  
Db 99 VAAAPVFAVSTSSAAGVAPNPAGSGNNSSPSSTSS-SSSSPSGSSLAESPEAA 157  
QY 65 KSETTIT-----QSGYNGADVQGADNSTIETQNG--FRNNATIDOWNAKNSDI 113  
Db 158 GVSSTAPLPGAGAGGTGVPVAGSALRELLEACRNGDVSRVKRLVDANVNAKDM 212

RESULT 4  
US-09-841-835-8  
; Sequence 8, Application US/09841835  
; Patent No. 6506587  
; GENERAL INFORMATION:  
; APPLICANT: de Lange, Titia  
; APPLICANT: Smith, Susan  
; TITLE OF INVENTION: A PROTEIN THAT BINDS TO TRF1 AND METHODS  
; TITLE OF INVENTION: OF USE THEREOF  
; NUMBER OF SEQUENCES: 12  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Klauber & Jackson  
; STREET: 411 Hackensack Avenue, 4th Floor  
; CITY: Hackensack  
; STATE: New Jersey  
; COUNTRY: USA  
; ZIP: 07601  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible

6 VAAFAAI-VVSGSALAGVVPOWGGGHNHGGNSSGPDSTLSIYQYGSANAALALQSDAR 64

Db 99 VAAAPVPAVTSAAAGVAPNPAGSGNNPSSSSPTSS-SSSSPSPGSSLAESPEAA 157  
QY 65 KSETTIT----QSGYNGADVGQADNSITELTONG--FRNNATIDOWNAKNSDI 113  
Db 158 GVSSTAPLGGAAGPGTGVPVAVSGALRELLACRNGDVSVRKRLVDAANVNAKDM 212

RESULT 7  
US-09-196-387-2  
; Sequence 2, Application US/09196387  
; Patent No. 6277613  
; GENERAL INFORMATION:  
; APPLICANT: de Lange, Titia  
; TITLE OF INVENTION: A PROTEIN THAT BINDS TO TRF1 AND METHODS  
; NUMBER OF SEQUENCES: 12  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Klauber & Jackson  
; STREET: 411 Hackensack Avenue, 4th Floor  
; CITY: Hackensack  
; STATE: New Jersey  
; COUNTRY: USA  
; ZIP: 07601  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; FILING DATE: US/09/196,387  
; CLASSIFICATION:  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: 09/095,225  
; FILING DATE: June 10, 1998  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Jackson Esq., David A.  
; REGISTRATION NUMBER: 26,742  
; REFERENCE/DOCKET NUMBER: 600-1-230 CIP1  
; TELEPHONE: 201-487-5800  
; TELEFAX: 201-343-1684  
; TELEX: 133521  
; INFORMATION FOR SEQ ID NO: 2:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 1327 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
; HYPOTHETICAL: NO  
; US-09-196-387-2

Query Match 11.9%; Score 92; DB 3; Length 1327;  
Best Local Similarity 30.4%; Pred. No. 1.3;  
Matches 35; Conservative 15; Mismatches 57; Indels 8; Gaps 4;  
QY 6 VAAAPAI-VVSGSALAGVVPQWGGNGHNGSGDPDSTLSIYQGSANAALQSDAR 64  
Db 99 VAAAPVPAVTSAAAGVAPNPAGSGNNPSSSSPTSS-SSSSPSPGSSLAESPEAA 157  
QY 65 KSETTIT----QSGYNGADVGQADNSITELTONG--FRNNATIDOWNAKNSDI 113  
Db 158 GVSSTAPLGGAAGPGTGVPVAVSGALRELLACRNGDVSVRKRLVDAANVNAKDM 212

RESULT 8  
US-09-841-835-2  
; Sequence 2, Application US/09841835  
; Patent No. 6506587  
; GENERAL INFORMATION:

; APPLICANT: de Lange, Titia  
; APPLICANT: Smith, Susan  
; TITLE OF INVENTION: A PROTEIN THAT BINDS TO TRF1 AND METHODS  
; NUMBER OF SEQUENCES: 12  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Klauber & Jackson  
; STREET: 411 Hackensack Avenue, 4th Floor  
; CITY: Hackensack  
; STATE: New Jersey  
; COUNTRY: USA  
; ZIP: 07601  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/09/841,835  
; FILING DATE:  
; CLASSIFICATION:  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: 09/196,387  
; FILING DATE:  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Jackson Esq., David A.  
; REGISTRATION NUMBER: 26,742  
; REFERENCE/DOCKET NUMBER: 600-1-230 CIP1  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 201-487-5800  
; TELEFAX: 201-343-1684  
; TELEX: 133521  
; INFORMATION FOR SEQ ID NO: 2:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 1327 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
; HYPOTHETICAL: NO  
; US-09-841-835-2

Query Match 11.9%; Score 92; DB 4; Length 1327;  
Best Local Similarity 30.4%; Pred. No. 1.3;  
Matches 35; Conservative 15; Mismatches 57; Indels 8; Gaps 4;

QY 6 VAAAPAI-VVSGSALAGVVPQWGGNGHNGSGDPDSTLSIYQGSANAALQSDAR 64  
Db 99 VAAAPVPAVTSAAAGVAPNPAGSGNNPSSSSPTSS-SSSSPSPGSSLAESPEAA 157  
QY 65 KSETTIT----QSGYNGADVGQADNSITELTONG--FRNNATIDOWNAKNSDI 113  
Db 158 GVSSTAPLGGAAGPGTGVPVAVSGALRELLACRNGDVSVRKRLVDAANVNAKDM 212

RESULT 9  
US-09-972-115A-8  
; Sequence 8, Application US/09972115A  
; Patent No. 6599728  
; GENERAL INFORMATION:  
; APPLICANT: Geron Corporation  
; APPLICANT: Gregg, Morin B.  
; APPLICANT: Walter, Funk D.  
; APPLICANT: Mieczyslaw, Platyszek A.  
; TITLE OF INVENTION: A Second Mammalian Telomerase  
; FILE REFERENCE: 080/003C  
; CURRENT APPLICATION NUMBER: US/09/972,115A  
; CURRENT FILING DATE: 2001-10-05  
; PRIOR APPLICATION NUMBER: US 60/128,577  
; PRIOR FILING DATE: 2000-04-10  
; PRIOR APPLICATION NUMBER: US 60/129,123  
; PRIOR FILING DATE: 1999-04-13  
; NUMBER OF SEQ ID NOS: 64



SOFTWARE: PatentIn version 3.1  
SEQ ID NO 8  
LENGTH: 1327  
TYPE: PRT  
ORGANISM: Homo sapiens  
US-09-972-115A-8

Query Match 11.9%; Score 92; DB 4; Length 1327;  
Best Local Similarity 30.4%; Pred. No. 1.3; Indels 8; Gaps 4;  
Matches 35; Conservative 15; Mismatches 57

QY 6 VAAFAAI-VVGSALAGVVPQWGGGNGHNGGNSGDPSTLSIYQYGSANAALALQSDAR 64  
Db 99 VAAAPVPAVTSAGVAPNAGSGNSPSSSSPTSS-SSSSSPSPGSSLAESPEAA 157  
QY 65 KSETTIT-OSQYNGADVGGADNSTIELTONG--FNNATIDOWNAKSDI 113  
Db 158 GVSSTAPLPGGAAGPOTGVPVAVSGALRELLACRNGDVSRKRLVDAANVAKDM 212

RESULT 10  
US-08-864-038A-3  
Sequence 3, Application US/08864038A  
Patent No. 6001592  
GENERAL INFORMATION:  
APPLICANT: KUNIO NAKASHIMA et al.  
TITLE OF INVENTION: NOVEL POLYPEPTIDE GENE CDNA, VECTOR  
TITLE OF INVENTION: CONTAINING SAID CDNA, HOST CELLS TRANSFORMED WITH SAID  
TITLE OF INVENTION: VECTOR, POLYPEPTIDE PRODUCED THEREBY, METHOD OF PRODUCING  
TITLE OF INVENTION: SAID POLYPEPTIDE, DNA ENCODING SAID POLYPEPTIDE AND ANTIBODY  
TITLE OF INVENTION: TO SAID POLYPEPTIDE  
NUMBER OF SEQUENCES: 4  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: 812-5 Hirano  
STREET: Isshinden  
CITY: Tsu-city  
STATE: Mie-prefecture  
COUNTRY: JAPAN  
ZIP: 514-01  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette, 3.50 inch, 1.44 MB storage  
COMPUTER: IBM Compatible  
OPERATING SYSTEM: Microsoft Windows 95  
SOFTWARE: Word Perfect 6.1  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/864,038A  
FILING DATE: May 28, 1997  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: JP 8-184459  
FILING DATE: 15-July-1996  
ATTORNEY/AGENT INFORMATION:  
NAME: C. Bruce Hamburg  
REGISTRATION NUMBER: 22,389  
REFERENCE/DOCKET NUMBER: F-5610  
TELEPHONE: (212)986-2340  
TELEFAX: (212)953-7733  
INFORMATION FOR SEQ ID NO: 3:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 738  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
ORIGINAL SOURCE:  
ORGANISM: Pinctada fucata  
CELL TYPE: mantle epithelial cell  
FEATURE:  
NAME/KEY: peptide  
LOCATION: from 1 to 738  
IDENTIFICATION METHOD: E (by experiment)  
US-08-864-038A-3

Query Match 11.5%; Score 89.5; DB 3; Length 738;

Best Local Similarity 35.4%; Pred. No. 1.1;  
Matches 29; Conservative 4; Mismatches 26; Indels 23; Gaps 3;  
QY 3 LLKVAFAAIVVGSALAGVVPQWGGGNGHNGGNSGDPSTLSIYQYGSANAALALQSD 62  
Db 419 LLKSSASASASASASAG-----GGGGGNGGNGGGG-----GGGAGALA----- 460  
QY 63 ARKSETTITQSGYNGADVGG 84  
Db 461 -----ALAAAGAGGLGGGG 477

RESULT 11  
US-08-968-685A-10  
Sequence 10, Application US/08968685A  
Patent No. 6214981  
GENERAL INFORMATION:  
APPLICANT: TUCKER, KENNETH  
TITLE OF INVENTION: FLOSIOLA, LAURA  
TITLE OF INVENTION: MORAXELLA CATARRHALIS OUTER MEMBRANE  
TITLE OF INVENTION: PROTEIN-106 POLYPEPTIDE, GENE SEQUENCE AND USES THEREOF  
NUMBER OF SEQUENCES: 21  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: PENNIE & EDMONDS LLP  
STREET: 1155 Avenue of the Americas  
CITY: New York  
STATE: New York  
COUNTRY: USA  
ZIP: 10036-2711  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/968,685A  
FILING DATE: No. 6214981ember 12, 1997  
CLASSIFICATION:  
ATTORNEY/AGENT INFORMATION:  
NAME: Balgwin, Geraldine F.  
REGISTRATION NUMBER: 31,232  
REFERENCE/DOCKET NUMBER: 7969-060  
TELEPHONE: (212) 790-9090  
TELEFAX: (212) 869-8864  
INFORMATION FOR SEQ ID NO: 10:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 2123 amino acids  
TYPE: amino acid  
STRANDEDNESS:  
TOPOLOGY: unknown  
MOLECULE TYPE: peptide  
US-08-968-685A-10

Query Match 11.2%; Score 87; DB 3; Length 2123;  
Best Local Similarity 25.4%; Pred. No. 7.6;  
Matches 48; Conservative 27; Mismatches 62; Indels 52; Gaps 11;

QY 1 MKLLKVAFAAIVV-----SGSALAGVVPQWGGGNGHNGGNSGDPSTLSIYQYGSANAA 56  
Db 44 LSFARIALAVLIVIGATLNGSAYAGIGISEADGG--KGGANARG-DKSIAT---GDIAQA 97  
QY 57 LALQSDARKSETTITQSGYNGADVGG--OGADNSTI-----ELTQNGFRNNA--TIDOWNAK 109  
Db 98 LGSQSIAGDKNKIVHNS--NNNANIGAKASNESIATGCDVLIAGSHASIAIGSDDLYLKK 155  
QY 110 NSDITVGYDQLVTRV-----VTHEMAHASVMVRQV----- 140  
Db 156 E---IVQQISELLPIIRGOKALNDIYQLADTNLQKRYRTHAQGHASTAVGAMSVAKGHFS 212  
QY 141 -GFGNNATA 148

Db 213 NAFGRATA 221

RESULT 12

US-09-252-991A-32096

Sequence 32096, Application US/09252991A

Patent No. 6551795

GENERAL INFORMATION:

APPLICANT: Marc J. Rubenfield et al.

TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS

FILE REFERENCE: 107196.136

CURRENT APPLICATION NUMBER: US/09/252,991A

CURRENT FILING DATE: 1999-02-18

PRIOR APPLICATION NUMBER: US 60/074,788

PRIOR FILING DATE: 1998-02-18

PRIOR APPLICATION NUMBER: US 60/094,190

PRIOR FILING DATE: 1998-07-27

NUMBER OF SEQ ID NOS: 33142

SEQ ID NO 32096

LENGTH: 339

TYPE: PRT

ORGANISM: Pseudomonas aeruginosa

US-09-252-991A-32096

Query Match 10.8%; Score 83.5; DB 4; Length 339;

Best Local Similarity 21.5%; Pred. No. 1.6;

Matches 35; Conservative 28; Mismatches 53; Indels 47; Gaps 7;

Qy 5 KVAAPAAIVVSGSALAGVVPQWGG-----GGNHGNGSGGPDSTLSIYQVGSANAAL 57

Db 58 KVNFGTL--NNASVSGSINKDASGVNVVAGDNNQQAALASADASFVFGTATAS- 114

Qy 58 ALODARKSTTTTOSQYNGADVGGADNSTLTQNGFRNATIDQNAKNS-----111

Db 115 -----TSVLQSYGN-----TLNYSNENTASLNSANNVSGNLG 149

Qy 112 -DITVGYDQ-----LVTRVVTHEMAHASVNVVRQVGGNNATAN 149

Db 150 VNVAGNFNQKNDLAAAVSNGQVSTAGSAASTGTG-NTTVN 191

RESULT 13

US-09-056-556-204

Sequence 204, Application US/09056556

Patent No. 6350456

GENERAL INFORMATION:

APPLICANT: Reed, Steven G.

APPLICANT: Skeiky, Yasir A.W.

APPLICANT: Dillon, Davin C.

TITLE OF INVENTION: COMPOUNDS AND METHODS FOR THE PREVENTION AND

NUMBER OF SEQUENCES: 241

CORRESPONDENCE ADDRESS:

ADDRESSEE: SEED and BERRY LLP

STREET: 6300 Columbia Center, 701 Fifth Avenue

CITY: Seattle

STATE: Washington

COUNTRY: USA

ZIP: 98104-7092

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: PatentIn Release #1.0, Version #1.30

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/09/056,556

FILING DATE: 07-APR-1998

CLASSIFICATION:

ATTORNEY/AGENT INFORMATION:

NAME: Maki, David J.

REGISTRATION NUMBER: 31,392

REFERENCE/DOCKET NUMBER: 210121.457

TELECOMMUNICATION INFORMATION:

TELEPHONE: (206) 622-4900

TELEFAX: (206) 682-6031

INFORMATION FOR SEQ ID NO: 199:

SEQUENCE CHARACTERISTICS:

LENGTH: 943 amino acids

TYPE: amino acid

STRANDEDNESS:

TOPOLOGY: linear

US-09-072-596-199

Query Match 10.8%; Score 83.5; DB 4; Length 943;

Best Local Similarity 25.4%; Pred. No. 6;

Matches 29; Conservative 16; Mismatches 50; Indels 19; Gaps 4;

TELEPHONE: (206) 622-4900

TELEFAX: (206) 682-6031

INFORMATION FOR SEQ ID NO: 204:

SEQUENCE CHARACTERISTICS:

LENGTH: 943 amino acids

TYPE: amino acid

STRANDEDNESS:

TOPOLOGY: linear

US-09-056-556-204

Query Match 10.8%; Score 83.5; DB 4; Length 943;

Best Local Similarity 25.4%; Pred. No. 6;

Matches 29; Conservative 16; Mismatches 50; Indels 19; Gaps 4;

Qy 12 IVVSGSALAGVVPQWGGGNGHNGSGGPDSTLSIYQVGSANAALQSDAR---KSET 68

Db 159 IGLTSGGLGF-----GGLNSGTGN-----IGLFNSGTGNVGNISGTGNWGIGNSG 205

Qy 69 TITQSYNGADVGGADNSTLTQNGFRNATIDQNAKNSD---ITVGYD 119

Db 206 NSYNTGFGSGDANTGFFNSGIANTGVGNAGNTYNTGSGNFGNSNTGFGNMGQYN 259

RESULT 14

US-09-072-596-199

Sequence 199, Application US/09072596

Patent No. 6458366

GENERAL INFORMATION:

APPLICANT: Reed, Steven G.

APPLICANT: Skeiky, Yasir A.W.

APPLICANT: Dillon, Davin C.

APPLICANT: Campos-Neto, Antonia

APPLICANT: Houghton, Raymond

APPLICANT: Wedvick, Thomas S.

APPLICANT: Twardzik, Daniel R.

APPLICANT: Lodes, Michael J.

APPLICANT: Hendrickson, Ronald C.

TITLE OF INVENTION: COMPOUNDS AND METHODS FOR DIAGNOSIS OF

NUMBER OF SEQUENCES: 350

CORRESPONDENCE ADDRESS:

ADDRESSEE: SEED and BERRY LLP

STREET: 6300 Columbia Center, 701 Fifth Avenue

CITY: Seattle

STATE: Washington

COUNTRY: USA

ZIP: 98104-7092

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: PatentIn Release #1.0, Version #1.30

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/09/072,596

FILING DATE: 05-MAY-1998

CLASSIFICATION:

ATTORNEY/AGENT INFORMATION:

NAME: Maki, David J.

REGISTRATION NUMBER: 31,392

REFERENCE/DOCKET NUMBER: 210121.417C9

TELECOMMUNICATION INFORMATION:

TELEPHONE: (206) 622-4900

TELEFAX: (206) 682-6031

INFORMATION FOR SEQ ID NO: 199:

SEQUENCE CHARACTERISTICS:

LENGTH: 943 amino acids

TYPE: amino acid

STRANDEDNESS:

TOPOLOGY: linear

US-09-072-596-199

Query Match 10.8%; Score 83.5; DB 4; Length 943;

Best Local Similarity 25.4%; Pred. No. 6;

Matches 29; Conservative 16; Mismatches 50; Indels 19; Gaps 4;

QY 12 IVVSGSALAGVWPQGGGNGHNGGSSGPDSTLSIYQYGSANAALALOSDAR---KSET 68  
Db 159 IGLTGSLLGF-----GGLNSGTGN-----IGLFNSGTGNVIGNSGTGNWGIGNSG 205  
QY 69 TITQSGYGNADYVQGADNSTIELTQNGFERNATIDQWNAKNSD---ITVGOYD 119  
Db 206 NSYNTGFGSGDANTGFFNSGIANTGVGNAGNYNTGSPNGNSNTGFGFNMGOYN 259

RESULT 15

US-09-477-135A-131  
; Sequence 131, Application US/09477135A  
; Patent No. 6572865  
; GENERAL INFORMATION:  
; APPLICANT: Nano, Francis  
; TITLE OF INVENTION: Mycobacterium Tuberculosis DNA Sequences Encoding  
; TITLE OF INVENTION: Immunostimulatory Peptides  
; FILE REFERENCE: 52888  
; CURRENT APPLICATION NUMBER: US/09/477,135A  
; CURRENT FILING DATE: 2000-01-03  
; PRIOR APPLICATION NUMBER: 08990823  
; PRIOR FILING DATE: 1997-12-15  
; PRIOR APPLICATION NUMBER: US 96/10375  
; PRIOR FILING DATE: 1996-06-14  
; PRIOR APPLICATION NUMBER: 60/000,254  
; PRIOR FILING DATE: 1995-06-15  
; NUMBER OF SEQ ID NOS: 169  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 131  
; LENGTH: 943  
; TYPE: PRT  
; ORGANISM: Mycobacterium tuberculosis  
US-09-477-135A-131

Query Match 10.8%; Score 83.5; DB 4; Length 943;  
Best Local Similarity 25.4%; Pred. No. 6;  
Matches 29; Conservative 16; Mismatches 50; Indels 19; Gaps 4;  
QY 12 IVVSGSALAGVWPQGGGNGHNGGSSGPDSTLSIYQYGSANAALALOSDAR---KSET 68  
Db 266 IGLTGSLLGF-----GGLNSGTGN-----IGLFNSGTGNVIGNSGTGNWGIGNSG 312  
QY 69 TITQSGYGNADYVQGADNSTIELTQNGFERNATIDQWNAKNSD---ITVGOYD 119  
Db 313 NSYNTGFGSGDANTGFFNSGIANTGVGNAGNYNTGSPNGNSNTGFGFNMGOYN 366

Search completed: August 2, 2004, 14:58:32  
Job time : 13 secs

This Page Blank (uspto)

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: August 2, 2004, 14:54:48 ; Search time 36.8 Seconds

(without alignments)  
1287.123 Million cell updates/sec

Title: US-09-543-407-14

Perfect score: 775

Sequence: 1 MKLLKVAFAAIVVSGSALA.....HASVMVQVFGNNATANYQ 151

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1291235 seqs, 313682936 residues

Total number of hits satisfying chosen parameters: 1291235

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published Applications AA:\*

1: /cgn2\_6/ptodata/2/pubpaa/US07\_PUBCOMB.pep.\*  
2: /cgn2\_6/ptodata/2/pubpaa/PCT\_NEW\_PUB.pep.\*  
3: /cgn2\_6/ptodata/2/pubpaa/US06\_NEW\_PUB.pep.\*  
4: /cgn2\_6/ptodata/2/pubpaa/US06\_PUBCOMB.pep.\*  
5: /cgn2\_6/ptodata/2/pubpaa/US07\_NEW\_PUB.pep.\*  
6: /cgn2\_6/ptodata/2/pubpaa/PCTUS\_PUBCOMB.pep.\*  
7: /cgn2\_6/ptodata/2/pubpaa/US08\_NEW\_PUB.pep.\*  
8: /cgn2\_6/ptodata/2/pubpaa/US08\_PUBCOMB.pep.\*  
9: /cgn2\_6/ptodata/2/pubpaa/US09A\_PUBCOMB.pep.\*  
10: /cgn2\_6/ptodata/2/pubpaa/US09B\_PUBCOMB.pep.\*  
11: /cgn2\_6/ptodata/2/pubpaa/US09C\_PUBCOMB.pep.\*  
12: /cgn2\_6/ptodata/2/pubpaa/US09D\_PUBCOMB.pep.\*  
13: /cgn2\_6/ptodata/2/pubpaa/US10A\_PUBCOMB.pep.\*  
14: /cgn2\_6/ptodata/2/pubpaa/US10B\_PUBCOMB.pep.\*  
15: /cgn2\_6/ptodata/2/pubpaa/US10C\_PUBCOMB.pep.\*  
16: /cgn2\_6/ptodata/2/pubpaa/US10D\_PUBCOMB.pep.\*  
17: /cgn2\_6/ptodata/2/pubpaa/US60\_NEW\_PUB.pep.\*  
18: /cgn2\_6/ptodata/2/pubpaa/US60\_PUBCOMB.pep.\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	525	67.7	151	12	US-09-741-873B-4
2	525	67.7	151	12	US-09-741-873B-4
3	447	57.7	131	12	US-09-741-873B-2
4	447	57.7	131	12	US-09-741-873B-2
5	104	13.4	445	15	US-10-369-493-20638
6	98	12.6	210	12	US-10-425-114-53421
7	98	12.6	211	12	US-10-425-114-53421
8	95	12.3	186	12	US-10-282-122A-49412
9	92	11.9	673	9	US-09-841-835-8
10	92	11.9	949	9	US-09-841-835-10
11	92	11.9	1327	9	US-09-841-835-2
12	92	11.9	1327	10	US-09-972-115A-8
13	92	11.9	1327	14	US-10-199-937-4
14	90.5	11.7	263	12	US-10-425-114-49960
15	88.5	11.4	145	16	US-10-437-963-147748

16	88	11.4	369	12	US-10-425-114-56041	Sequence 56041, A
17	88	11.4	486	12	US-10-424-599-275468	Sequence 275468, A
18	88	11.4	507	12	US-10-425-114-57763	Sequence 57763, A
19	87	11.2	597	9	US-09-793-306-146	Sequence 146, App
20	87	11.2	628	12	US-10-282-122A-53269	Sequence 53269, A
21	87	11.2	678	12	US-10-282-122A-64573	Sequence 64573, A
22	87	11.2	2122	9	US-09-813-214A-9	Sequence 9, Appli
23	86.5	11.2	527	9	US-09-712-363-156	Sequence 156, App
24	86.5	11.2	527	14	US-10-080-170-348	Sequence 348, App
25	86.5	11.2	527	16	US-10-080-170-348	Sequence 348, App
26	86	11.1	482	14	US-10-156-761-8763	Sequence 8763, Ap
27	86	11.1	486	15	US-10-369-493-20619	Sequence 20619, A
28	85.5	11.0	270	16	US-10-437-963-122263	Sequence 122263, A
29	85.5	11.0	1048	14	US-10-174-363-56	Sequence 56, Appl
30	85.5	11.0	1048	15	US-10-374-780A-2086	Sequence 2086, Ap
31	85	11.0	438	14	US-10-156-761-9343	Sequence 9343, Ap
32	84.5	10.9	271	14	US-10-156-761-11721	Sequence 11721, A
33	83.5	10.8	943	9	US-09-996-634-131	Sequence 131, App
34	83.5	10.8	943	10	US-09-997-182-131	Sequence 131, App
35	83.5	10.8	943	10	US-09-997-181-131	Sequence 131, App
36	83.5	10.8	943	14	US-10-133-002-199	Sequence 199, App
37	83.5	10.8	943	14	US-10-084-843-204	Sequence 204, App
38	83.5	10.8	3300	12	US-10-282-122A-64369	Sequence 64369, A
39	83	10.7	242	12	US-10-425-114-61520	Sequence 61520, A
40	83	10.7	254	10	US-09-880-748-1136	Sequence 1136, Ap
41	83	10.7	254	10	US-09-880-748-1165	Sequence 1165, Ap
42	83	10.7	254	12	US-10-293-418-1136	Sequence 1136, Ap
43	83	10.7	254	12	US-10-293-418-1165	Sequence 1165, Ap
44	83	10.7	257	10	US-09-880-748-1494	Sequence 1494, Ap
45	83	10.7	257	12	US-10-293-418-1494	Sequence 1494, Ap

#### ALIGNMENTS

#### RESULT 1

US-09-741-873B-4  
; Sequence 4, Application US/09741873B  
; Publication No. US20020081722A1  
; GENERAL INFORMATION:  
; APPLICANT: Normark, Staffan  
; APPLICANT: Olsen, Arne  
; TITLE OF INVENTION: Fibronectin Binding Protein As Well As Its Preparation  
; FILE REFERENCE: 012889-084  
; CURRENT APPLICATION NUMBER: US/09/741,873B  
; CURRENT FILING DATE: 2003-04-04  
; PRIOR APPLICATION NUMBER: SE 8801723-1  
; PRIOR FILING DATE: 1998-05-06  
; PRIOR APPLICATION NUMBER: US 08/978,878  
; PRIOR FILING DATE: 1997-11-26  
; PRIOR APPLICATION NUMBER: US 07/347,189  
; PRIOR FILING DATE: 1989-05-04  
; PRIOR APPLICATION NUMBER: US 07/789,437  
; PRIOR FILING DATE: 1991-11-06  
; PRIOR APPLICATION NUMBER: US 07/970,846  
; PRIOR FILING DATE: 1992-11-03  
; PRIOR APPLICATION NUMBER: US 08/187,865  
; PRIOR FILING DATE: 1994-01-28  
; PRIOR APPLICATION NUMBER: US 08/318,519  
; NUMBER OF SEQ ID NOS: 10  
; SOFTWARE: PatentIn version 3.0  
; SEQ ID NO 4  
; LENGTH: 151  
; TYPE: PRT  
; ORGANISM: Escherichia coli  
US-09-741-873B-4

Query Match 67.7%; Score 525; DB 12; Length 151;  
Best Local Similarity 68.2%; Pred. No. 2.1e-45;  
Matches 103; Conservative 21; Mismatches 27; Indels 0; Gaps 0;

QY 1 MKLLKVAFAAIVVSGSALAGVVPQWGGGNGHNGSSGPDSTLSTIYQGSANAALQ 60

```

Db 1 MLLKVAATAAIVFSSAVGVVPOYGGGNGGNGNSGPNSELNIYQYGGNSALALQ 60
    :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Qy 61 SDARKSETTITQSGYNGADVGCGADNSTIELTQNGFRNATIDQWNAKNSDITVQGYDQ 120
    :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 61 TDARNSDLTITQHGCGNGADVGCGSDSSIDLTLQRFNGSATLDQWNGKNSMTVKQFGG 120
    :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Qy 121 LVTRVVTHEMAHASVVMVQVQFGNNATANQY 151
    :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 121 NGGAADVDTASNSVNVTVQVGFNNATAHQY 151
    :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:

```

```

RESULT 2
US-09-741-873B-4
; Sequence 4, Application US/09741873B
; Publication No. US2004009695A9
; GENERAL INFORMATION:
; APPLICANT: Normark, Staffan
; TITLE OF INVENTION: Fibronectin Binding Protein As Well As Its Preparation
; FILE REFERENCE: 012889-084
; CURRENT APPLICATION NUMBER: US/09/741,873B
; PRIOR FILING DATE: 2003-04-04
; PRIOR APPLICATION NUMBER: SE 8801723-1
; PRIOR FILING DATE: 1998-05-06
; PRIOR APPLICATION NUMBER: US 08/978,878
; PRIOR FILING DATE: 1997-11-26
; PRIOR APPLICATION NUMBER: US 07/347,189
; PRIOR FILING DATE: 1989-05-04
; PRIOR APPLICATION NUMBER: US 07/789,437
; PRIOR FILING DATE: 1991-11-06
; PRIOR APPLICATION NUMBER: US 07/970,846
; PRIOR FILING DATE: 1992-11-03
; PRIOR APPLICATION NUMBER: US 08/187,865
; PRIOR FILING DATE: 1994-01-28
; PRIOR APPLICATION NUMBER: US 08/318,519
; PRIOR FILING DATE: 1994-10-05
; NUMBER OF SEQ ID NOS: 10
; SOFTWARE: Patent in version 3.0
; SEQ ID NO 4
; LENGTH: 151
; TYPE: PRT
; ORGANISM: Escherichia coli
US-09-741-873B-4

```

```

Query Match 67.7%; Score 525; DB 12; Length 151;
Best Local Similarity 68.2%; Pred. No. 2,le-45;
Matches 103; Conservative 21; Mismatches 27; Indels 0; Gaps 0;

Qy 1 MLLKVAATAAIVFSSALAGVVPQVGGGNGGNGNSGPNSELNIYQYGSANAALALQ 60
    :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 1 MLLKVAATAAIVFSSAGAVGVVPOYGGGNGGNGNSGPNSELNIYQYGGNSALALQ 60
    :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Qy 61 SDARKSETTITQSGYNGADVGCGADNSTIELTQNGFRNATIDQWNAKNSDITVQGYDQ 120
    :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 61 TDARNSDLTITQHGCGNGADVGCGSDSSIDLTLQRFNGSATLDQWNGKNSMTVKQFGG 120
    :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Qy 121 LVTRVVTHEMAHASVVMVQVQFGNNATANQY 151
    :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 121 NGGAADVDTASNSVNVTVQVGFNNATAHQY 151
    :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:

```

```

RESULT 3
US-09-741-873B-2
; Sequence 2, Application US/09741873B
; Publication No. US20020081722A1
; GENERAL INFORMATION:
; APPLICANT: Normark, Staffan
; TITLE OF INVENTION: Fibronectin Binding Protein As Well As Its Preparation
; FILE REFERENCE: 012889-084
; CURRENT APPLICATION NUMBER: US/09/741,873B
; PRIOR FILING DATE: 2003-04-04

```

```

; PRIOR APPLICATION NUMBER: SE 8801723-1
; PRIOR FILING DATE: 1998-05-06
; PRIOR APPLICATION NUMBER: US 08/978,878
; PRIOR FILING DATE: 1997-11-26
; PRIOR APPLICATION NUMBER: US 07/347,189
; PRIOR FILING DATE: 1989-05-04
; PRIOR APPLICATION NUMBER: US 07/789,437
; PRIOR FILING DATE: 1991-11-06
; PRIOR APPLICATION NUMBER: US 07/970,846
; PRIOR FILING DATE: 1992-11-03
; PRIOR APPLICATION NUMBER: US 08/187,865
; PRIOR FILING DATE: 1994-01-28
; PRIOR APPLICATION NUMBER: US 08/318,519
; PRIOR FILING DATE: 1994-10-05
; NUMBER OF SEQ ID NOS: 10
; SOFTWARE: Patent in version 3.0
; SEQ ID NO 2
; LENGTH: 131
; TYPE: PRT
; ORGANISM: Escherichia coli
US-09-741-873B-2

Query Match 57.7%; Score 447; DB 12; Length 131;
Best Local Similarity 64.9%; Pred. No. 1,5e-37;
Matches 85; Conservative 20; Mismatches 26; Indels 0; Gaps 0;

Qy 21 GVPQWGGGNGHNGGNSGPDSTLSIYQYGSANAALALQSDARKSETTITQSGYNGAD 80
    :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 1 GVPQYGGGNGHNGGNSGENSELTIYQYGGNSALALQTDARNSDLTITQHGCGNGAD 60
    :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Qy 81 VQGGADNSTIELTQNGFRNATIDQWNAKNSDITVQGYDQVTRVVTHEMAHASVVMVQV 140
    :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 61 VQGSDDSSIDLTLQRFNGSATLDQWNGKNSMTVKQFGCGGAADVDTASNSVNVTVQ 120
    :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Qy 141 GFGNNATANQY 151
    :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 121 GFGNNATAHQY 131
    :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:

```

```

RESULT 4
US-09-741-873B-2
; Sequence 2, Application US/09741873B
; Publication No. US2004009695A9
; GENERAL INFORMATION:
; APPLICANT: Normark, Staffan
; TITLE OF INVENTION: Fibronectin Binding Protein As Well As Its Preparation
; FILE REFERENCE: 012889-084
; CURRENT APPLICATION NUMBER: US/09/741,873B
; PRIOR FILING DATE: 2003-04-04
; PRIOR APPLICATION NUMBER: SE 8801723-1
; PRIOR FILING DATE: 1998-05-06
; PRIOR APPLICATION NUMBER: US 08/978,878
; PRIOR FILING DATE: 1997-11-26
; PRIOR APPLICATION NUMBER: US 07/347,189
; PRIOR FILING DATE: 1989-05-04
; PRIOR APPLICATION NUMBER: US 07/789,437
; PRIOR FILING DATE: 1991-11-06
; PRIOR APPLICATION NUMBER: US 07/970,846
; PRIOR FILING DATE: 1992-11-03
; PRIOR APPLICATION NUMBER: US 08/187,865
; PRIOR FILING DATE: 1994-01-28
; PRIOR APPLICATION NUMBER: US 08/318,519
; PRIOR FILING DATE: 1994-10-05
; NUMBER OF SEQ ID NOS: 10
; SOFTWARE: Patent in version 3.0
; SEQ ID NO 2
; LENGTH: 131
; TYPE: PRT
; ORGANISM: Escherichia coli
US-09-741-873B-2

Query Match 57.7%; Score 447; DB 12; Length 131;

```

```
Best Local Similarity 64.9%; Pred. No. 1.5e-37;
Matches 85; Conservative 20; Mismatches 26; Indels 0; Gaps 0;
QY 21 GVVPQGGGNNHGGNSGPDSTLSIYQYGSANAALQSDARKSETTITQSGYNGAD 80
Db 1 GVVPQGGGNNHGGNSGPDSTLSIYQYGSANAALQSDARKSETTITQSGYNGAD 60
QY 81 VGGADNSTIELTQNGFRNATIDQWNAKNSDITVQYQDLVTRVVTTHMAHASVVRQV 140
Db 61 VGGSDSSIDLQRFSGNSATLDQWNAKNSDITVQYQDLVTRVVTTHMAHASVVRQV 120
QY 141 GFGNNATANQY 151
Db 121 GFGNNATANQY 131

RESULT 5
US-10-369-493-20638
; Sequence 20638, Application US/10369493
; Publication No. US20030233675A1
; GENERAL INFORMATION:
; APPLICANT: Cao, Yongwei
; APPLICANT: Hinkle, Gregory J.
; APPLICANT: Slater, Steven C.
; APPLICANT: Goldman, Barry S.
; APPLICANT: Chen, Xianfeng
; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
; FILE REFERENCE: 38-10(52052)B
; CURRENT APPLICATION NUMBER: US/10/369,493
; CURRENT FILING DATE: 2003-02-28
; PRIOR APPLICATION NUMBER: US 60/360,039
; PRIOR FILING DATE: 2002-02-21
; NUMBER OF SEQ ID NOS: 47374
; SEQ ID NO 20638
; LENGTH: 445
; TYPE: PRT
; ORGANISM: Rhodospseudomonas palustris
; FEATURE:
; NAME/KEY: unsure
; LOCATION: (1)..(445)
; OTHER INFORMATION: unsure at all Xaa locations
US-10-369-493-20638

Query Match 13.4%; Score 104; DB 15; Length 445;
Best Local Similarity 26.6%; Pred. No. 0.051;
Matches 45; Conservative 25; Mismatches 61; Indels 38; Gaps 8;
QY 7 AAFAA-----IVVSGSALAGVVPQGGGG-----NHNGG-----GNSGPDSTLSIYQY 50
Db 19 AFAADSNVYVLTQNGDQANITQSGNGNSVGAENGSGFLOENGLTSGA-NLLIVKQS 77
QY 51 GSANAALQSDARKSETTITQSGYNGADVGGADNSTIELTQNGFRNATIDQWNAK 110
Db 78 GNSNSV-----GRDIQKQSGAGNSAAIFQEGTSDVLEQQTGTSNGAVPSGWNWTN 129
QY 111 SDITVQYQDLVTRVVTTHMAHASV-----VRQVFGNNATANQ 150
Db 130 DP---GVFNK-ITQDSSNGSKSVIQQDKNNVFSIKQNTGNTSVNQ 174

RESULT 6
US-10-425-114-53421
; Sequence 53421, Application US/10425114
; Publication No. US20040034888A1
; GENERAL INFORMATION:
; APPLICANT: Liu, Jingdong
; APPLICANT: Zhou, Yihua
; APPLICANT: Kovalic, David K.
; APPLICANT: Screen, Steven E.
; APPLICANT: Tabaska, Jack E
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated with
```

```
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(53313)B
; CURRENT APPLICATION NUMBER: US/10/425,114
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 73128
; SEQ ID NO 53421
; LENGTH: 210
; TYPE: PRT
; ORGANISM: Glycine max
; FEATURE:
; OTHER INFORMATION: Clone ID: 700839445_FLI.pep
US-10-425-114-53421

Query Match 12.6%; Score 98; DB 12; Length 210;
Best Local Similarity 26.7%; Pred. No. 0.08;
Matches 36; Conservative 18; Mismatches 65; Indels 16; Gaps 4;
QY 26 WGGGNNHGGNSGPDSTLSIYQY-----GSANAALQSDARKSETTITQSGYNG 78
Db 61 WGGGGGGGGGGGSGASNVRAHYEPEQHGWLNAVSAYCSTWDASKPYWSKYGWT 120
QY 79 ADVG---QGADN--STIELTQNGFRNNA---TIDQWNAKNSDITVQYQDLVTRVVT 129
Db 121 AFCGPVGPGRDSCGKCLRVTNTGTGANTIVRIVDQCSNGGLDLVGVFNRIDTDGRGYQ 180
QY 130 MAHASVVRQVQFGN 144
Db 181 QGHLIVNYQFVDCGN 195

RESULT 7
US-10-424-599-245046
; Sequence 245046, Application US/10424599
; Publication No. US20040031072A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa Thomas J
; APPLICANT: Kovalic David K
; APPLICANT: Zhou Yihua
; APPLICANT: Cao Yongwei
; TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated with
; FILE REFERENCE: 38-21(53223)B
; CURRENT APPLICATION NUMBER: US/10/424,599
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 285684
; SEQ ID NO 245046
; LENGTH: 211
; TYPE: PRT
; ORGANISM: Glycine max
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT3847_63306C.1.pep
US-10-424-599-245046

Query Match 12.6%; Score 98; DB 12; Length 211;
Best Local Similarity 26.7%; Pred. No. 0.08;
Matches 36; Conservative 18; Mismatches 65; Indels 16; Gaps 4;
QY 26 WGGGNNHGGNSGPDSTLSIYQY-----GSANAALQSDARKSETTITQSGYNG 78
Db 62 WGGGGGGGGGGGSGASNVRAHYEPEQHGWLNAVSAYCSTWDASKPYWSKYGWT 121
QY 79 ADVG---QGADN--STIELTQNGFRNNA---TIDQWNAKNSDITVQYQDLVTRVVT 129
Db 122 AFCGPVGPGRDSCGKCLRVTNTGTGANTIVRIVDQCSNGGLDLVGVFNRIDTDGRGYQ 181
QY 130 MAHASVVRQVQFGN 144
Db 182 QGHLIVNYQFVDCGN 196

RESULT 8
US-10-282-122A-49412
; Sequence 49412, Application US/10282122A
```

```
Publication No. US20040029129A1
GENERAL INFORMATION:
APPLICANT: Wang, Liangsu
APPLICANT: Zamudio, Carlos
APPLICANT: Malone, Cheryl
APPLICANT: Haselbeck, Robert
APPLICANT: Ohlsen, Kari
APPLICANT: Zyskind, Judith
APPLICANT: Wall, Daniel
APPLICANT: Trawick, John
APPLICANT: Carr, Grant
APPLICANT: Yamamoto, Robert
APPLICANT: Forsyth, R.
APPLICANT: Xu, H.
TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
FILE REFERENCE: ELITRA.034A
CURRENT APPLICATION NUMBER: US/10/282,122A
CURRENT FILING DATE: 2003-02-20
PRIOR APPLICATION NUMBER: 60/191,078
PRIOR FILING DATE: 2000-03-21
PRIOR APPLICATION NUMBER: 60/206,848
PRIOR FILING DATE: 2000-05-23
PRIOR APPLICATION NUMBER: 60/207,727
PRIOR FILING DATE: 2000-05-26
PRIOR APPLICATION NUMBER: 60/230,335
PRIOR FILING DATE: 2000-09-06
PRIOR APPLICATION NUMBER: 60/230,347
PRIOR FILING DATE: 2000-09-09
PRIOR APPLICATION NUMBER: 60/242,578
PRIOR FILING DATE: 2000-10-23
PRIOR APPLICATION NUMBER: 60/253,625
PRIOR FILING DATE: 2000-11-27
PRIOR APPLICATION NUMBER: 60/257,931
PRIOR FILING DATE: 2000-12-22
PRIOR APPLICATION NUMBER: 60/267,636
PRIOR FILING DATE: 2001-02-09
PRIOR APPLICATION NUMBER: 60/269,308
PRIOR FILING DATE: 2001-02-16
Remaining Prior Application data removed - See File Wrapper or PALM.
NUMBER OF SEQ ID NOS: 78614
SOFTWARE: PatentIn version 3.1
SEQ ID NO 49412
LENGTH: 186
TYPE: PRT
ORGANISM: Burkholderia fungorum
US-10-282-122A-49412

Query Match 12.3%; Score 95; DB 12; Length 186;
Best Local Similarity 29.9%; Pred. No. 0.14;
Matches 50; Conservative 20; Mismatches 59; Indels 38; Gaps 11;

QY 1 MKLLKVAFAAIIVVSGSALAGVWPQGGGNGHNGGNS-----GPDSTLSIYQGSANAA 56
Db 1 MKL-----SLPAIAAASLGLVPL-----GAHADGTISITGVITDSTCSI--NGNANG- 49
QY 57 LALQSDARKSETTITQSGYNGADVQ--QADNST--IELTQNG-----FRNNT 102
Db 50 ----TPADKAITLATVPA--GSLASAGAVAGTSNPTDQLSLTCTGTATKAIARFENGPT 104
QY 103 IDQWNAKNSDITQYDQVLTATVTHMAHASVYVRQVGVGNATAN 149
Db 105 VDQNGVLSN--TAGTAQNEVRLINAQMQPINV---TTGANDDITN 147

RESULT 9
US-09-841-835-8
Sequence 8, Application US/09841835
Patent No. US20020076795A1
GENERAL INFORMATION:
APPLICANT: de Lange, Titia
APPLICANT: Smith, Susan
TITLE OF INVENTION: A PROTEIN THAT BINDS TO TRF1 AND METHODS
TITLE OF INVENTION: OF USE THEREOF

NUMBER OF SEQUENCES: 12
CORRESPONDENCE ADDRESS:
ADDRESS: Klauber & Jackson
STREET: 411 Hackensack Avenue, 4th Floor
CITY: Hackensack
STATE: New Jersey
COUNTRY: USA
ZIP: 07601
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/841,835
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 09/196,387
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Jackson Esq., David A.
REGISTRATION NUMBER: 26,742
REFERENCE/DOCKET NUMBER: 600-1-230 CIPI
TELECOMMUNICATION INFORMATION:
TELEPHONE: 201-487-5800
TELEFAX: 201-343-1684
TELEX: 133521
INFORMATION FOR SEQ ID NO: 8:
SEQUENCE CHARACTERISTICS:
LENGTH: 673 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-09-841-835-8

Query Match 11.9%; Score 92; DB 9; Length 673;
Best Local Similarity 30.4%; Pred. No. 1.4;
Matches 35; Conservative 15; Mismatches 57; Indels 8; Gaps 4;

QY 6 VAAFAAI-VVSGSALAGVWPQGGGNGHNGGNSGPDSTLSIYQGSANAAALQSDAR 64
Db 99 VAAAPVPAVSTSSAAGVAPNPAGSGNNSPSSSSSTSS--SSSSPSSPGSSLAESPAA 157
QY 65 KSETTIT-----QSGYNGADVQGDADNSTIELTQNG--FRNNTIDQWNAKNSDI 113
Db 158 GVSSTAPLPGAGAGTGTGVPVAVSGALRELEACRNGDVSRVTKLVDAANVNAKDM 212

RESULT 10
US-09-841-835-10
Sequence 10, Application US/09841835
Patent No. US20020076795A1
GENERAL INFORMATION:
APPLICANT: de Lange, Titia
APPLICANT: Smith, Susan
TITLE OF INVENTION: A PROTEIN THAT BINDS TO TRF1 AND METHODS
TITLE OF INVENTION: OF USE THEREOF
NUMBER OF SEQUENCES: 12
CORRESPONDENCE ADDRESS:
ADDRESS: Klauber & Jackson
STREET: 411 Hackensack Avenue, 4th Floor
CITY: Hackensack
STATE: New Jersey
COUNTRY: USA
ZIP: 07601
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/841,835
```



FILING DATE:  
CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 09/196,387  
FILING DATE:  
ATTORNEY/AGENT INFORMATION:  
NAME: Jackson Esq., David A.  
REGISTRATION NUMBER: 26,742  
REFERENCE/DOCKET NUMBER: 600-1-230 CIP1  
TELEPHONE: 201-487-5800  
TELEFAX: 201-343-1684  
TELEX: 133521  
INFORMATION FOR SEQ ID NO: 10:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 949 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-09-841-835-10

Query Match 11.9%; Score 92; DB 9; Length 949;  
Best Local Similarity 30.4%; Pred. No. 2.2; Mismatches 57; Indels 8; Gaps 4;  
Matches 35; Conservative 15;

QY 6 VAAFAAI-VVSGSALAGVVPQWGGGNNHGGNNSGDPSTLSIYQVGSANAALALQSDAR 64  
DB 99 VAAAPVPAVSTSSAAGVAPNPAGSGNNSPSSSSPTSS-SSSPSPSGSLAESPEAA 157  
QY 65 KSETTIT-----QSGYNGADVGQADNSTIELTONG--FRNNATIDOWNAKNSDI 113  
DB 158 GVSSTAPLPGAGPGTGVPAVSGALRELLACRNGDVSVRKRLVDAANVNAKDM 212

RESULT 11  
US-09-841-835-2  
Sequence 2, Application US/09841835  
Patent No. US20020076795A1  
GENERAL INFORMATION:  
APPLICANT: Smith, Susan  
TITLE OF INVENTION: A PROTEIN THAT BINDS TO TR1 AND METHODS  
TITLE OF INVENTION: OF USE THEREOF  
NUMBER OF SEQUENCES: 12  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Klauber & Jackson  
STREET: 411 Hackensack Avenue, 4th Floor  
CITY: Hackensack  
STATE: New Jersey  
COUNTRY: USA  
ZIP: 07601  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent In Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
FILING DATE: US/09/841,835  
CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 09/196,387  
FILING DATE:  
ATTORNEY/AGENT INFORMATION:  
NAME: Jackson Esq., David A.  
REGISTRATION NUMBER: 26,742  
REFERENCE/DOCKET NUMBER: 600-1-230 CIP1  
TELEPHONE: 201-487-5800  
TELEFAX: 201-343-1684  
TELEX: 133521  
INFORMATION FOR SEQ ID NO: 2:  
SEQUENCE CHARACTERISTICS:

FILING DATE:  
CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 09/196,387  
FILING DATE:  
ATTORNEY/AGENT INFORMATION:  
NAME: Jackson Esq., David A.  
REGISTRATION NUMBER: 26,742  
REFERENCE/DOCKET NUMBER: 600-1-230 CIP1  
TELEPHONE: 201-487-5800  
TELEFAX: 201-343-1684  
TELEX: 133521  
INFORMATION FOR SEQ ID NO: 2:  
SEQUENCE CHARACTERISTICS:

FILING DATE:  
CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 09/196,387  
FILING DATE:  
ATTORNEY/AGENT INFORMATION:  
NAME: Jackson Esq., David A.  
REGISTRATION NUMBER: 26,742  
REFERENCE/DOCKET NUMBER: 600-1-230 CIP1  
TELEPHONE: 201-487-5800  
TELEFAX: 201-343-1684  
TELEX: 133521  
INFORMATION FOR SEQ ID NO: 2:  
SEQUENCE CHARACTERISTICS:

LENGTH: 1327 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
HYPOTHETICAL: NO  
US-09-841-835-2

Query Match 11.9%; Score 92; DB 9; Length 1327;  
Best Local Similarity 30.4%; Pred. No. 3.3; Mismatches 57; Indels 8; Gaps 4;  
Matches 35; Conservative 15;

QY 6 VAAFAAI-VVSGSALAGVVPQWGGGNNHGGNNSGDPSTLSIYQVGSANAALALQSDAR 64  
DB 99 VAAAPVPAVSTSSAAGVAPNPAGSGNNSPSSSSPTSS-SSSPSPSGSLAESPEAA 157  
QY 65 KSETTIT-----QSGYNGADVGQADNSTIELTONG--FRNNATIDOWNAKNSDI 113  
DB 158 GVSSTAPLPGAGPGTGVPAVSGALRELLACRNGDVSVRKRLVDAANVNAKDM 212

RESULT 12  
US-09-972-115A-8  
Sequence 8, Application US/09972115A  
Publication No. US20030032769A1  
GENERAL INFORMATION:  
APPLICANT: Genon Corporation  
APPLICANT: Gregg, Morin B.  
APPLICANT: Walter, Funk D.  
APPLICANT: Misczyslaw, Piatyszek A.  
TITLE OF INVENTION: A Second Mammalian Telomerase  
FILE REFERENCE: 080/003C  
CURRENT APPLICATION NUMBER: US/09/972,115A  
CURRENT FILING DATE: 2001-10-05  
PRIOR APPLICATION NUMBER: US 60/128,577  
PRIOR FILING DATE: 2000-04-10  
PRIOR APPLICATION NUMBER: US 60/129,123  
PRIOR FILING DATE: 1999-04-13  
NUMBER OF SEQ ID NOS: 64  
SOFTWARE: Patent in version 3.1  
SEQ ID NO 8  
LENGTH: 1327  
TYPE: PRT  
ORGANISM: Homo sapiens  
US-09-972-115A-8

Query Match 11.9%; Score 92; DB 10; Length 1327;  
Best Local Similarity 30.4%; Pred. No. 3.3; Mismatches 57; Indels 8; Gaps 4;  
Matches 35; Conservative 15;

QY 6 VAAFAAI-VVSGSALAGVVPQWGGGNNHGGNNSGDPSTLSIYQVGSANAALALQSDAR 64  
DB 99 VAAAPVPAVSTSSAAGVAPNPAGSGNNSPSSSSPTSS-SSSPSPSGSLAESPEAA 157  
QY 65 KSETTIT-----QSGYNGADVGQADNSTIELTONG--FRNNATIDOWNAKNSDI 113  
DB 158 GVSSTAPLPGAGPGTGVPAVSGALRELLACRNGDVSVRKRLVDAANVNAKDM 212

RESULT 13  
US-10-199-937-4  
Sequence 4, Application US/10199937  
Publication No. US20030190739A1  
GENERAL INFORMATION:  
APPLICANT: Christenson, Erik  
APPLICANT: Demaggio, Anthony J.  
APPLICANT: Goldmat, Phyllis S.  
APPLICANT: Mcelligott, David L.  
TITLE OF INVENTION: TANKYRASE2 MATERIALS AND METHODS  
FILE REFERENCE: 27866/36559  
CURRENT APPLICATION NUMBER: US/10/199,937  
CURRENT FILING DATE: 2002-07-22  
PRIOR APPLICATION NUMBER: US/09/606,035

; PRIOR FILING DATE: 2000-06-28  
; PRIOR APPLICATION NUMBER: 60/141,582  
; PRIOR FILING DATE: 1999-06-29  
; NUMBER OF SEQ ID NOS: 178  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 4  
; LENGTH: 1327  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-10-199-937-4

Query Match  
Best Local Similarity 11.9%; Score 92; DB 14; Length 1327;  
Matches 35; Conservative 15; Mismatches 57; Indels 8; Gaps 4;

QY 6 VAAFAAI-VVSGSALAGVVPQWGGGNGHGGNSGPDSTLSIYQYGSNAALALQSDAR 64  
DB 99 VAAAPVPAVSTSSAAGVAPNPAGSGNSPSSSSPTSS-SSSFPSPGSSLAESPAA 157  
QY 65 KSETTIT---OSGYNGADVGOGADNSTIELTQNG--FRNNATIDQWNAKNSDI 113  
DB 158 GVSSTAPLPGAGPGTGPVAVSGALRELLACRNGDVSEVKELVDANVNAKM 212

RESULT 14  
US-10-425-114-49960  
; Sequence 49960, Application US/10425114  
; Publication No. US2004003488A1  
; GENERAL INFORMATION:  
; APPLICANT: Liu, Jingdong  
; APPLICANT: Zhou, Yihua  
; APPLICANT: Kovalic, David K.  
; APPLICANT: Screen, Steven E  
; APPLICANT: Tabaska, Jack E  
; APPLICANT: Cao, Yongwei  
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With  
; FILE REFERENCE: 38-21(53313)B  
; CURRENT APPLICATION NUMBER: US/10/425,114  
; CURRENT FILING DATE: 2003-04-28  
; NUMBER OF SEQ ID NOS: 73128  
; SEQ ID NO 49960  
; LENGTH: 263  
; TYPE: PRT  
; ORGANISM: Zea mays  
; FEATURE:  
; OTHER INFORMATION: Clone ID: 700071884\_FLI.pep  
US-10-425-114-49960

Query Match  
Best Local Similarity 11.7%; Score 90.5; DB 12; Length 263;  
Matches 29; Conservative 15; Mismatches 39; Indels 21; Gaps 4;

QY 17 SALAGVVPQWGGGNGHGGNSGPDSTLSIYQYGSNAALALQSDARKSETTITQSG-- 74  
DB 84 SSIAG-----GGGGGGGGGGTNGSGGSGGSGSSSTRAAGSPSGNYADAEKGAG 138  
QY 75 -----YNGA--DVCGADNSTIEL--TQNGFRNNATID 104  
DB 139 GGMGGGANGAYGSGAGGVGKGEVSGVALAPSSNGYNGGAAD 182

RESULT 15  
US-10-437-963-147748  
; Sequence 147748, Application US/10437963  
; Publication No. US20040123343A1  
; GENERAL INFORMATION:  
; APPLICANT: La Rosa, Thomas J.  
; APPLICANT: Kovalic, David K.  
; APPLICANT: Zhou, Yihua  
; APPLICANT: Cao, Yongwei  
; APPLICANT: Wu, Wei  
; APPLICANT: Boukharov, Andrey A.

; APPLICANT: Barbazuk, Brad  
; APPLICANT: Li, Ping  
; TITLE OF INVENTION: Rice Nucleic Acid Molecules and Other Molecules Associated With  
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement  
; FILE REFERENCE: 38-21(53221)B  
; CURRENT APPLICATION NUMBER: US/10/437,963  
; CURRENT FILING DATE: 2003-05-14  
; NUMBER OF SEQ ID NOS: 204966  
; SEQ ID NO 147748  
; LENGTH: 145  
; TYPE: PRT  
; ORGANISM: Oryza sativa  
; FEATURE:  
; OTHER INFORMATION: Clone ID: PAT\_MRT4530\_48247C.1.pep  
US-10-437-963-147748

Query Match  
Best Local Similarity 11.4%; Score 88.5; DB 16; Length 145;  
Matches 29; Conservative 11; Mismatches 36; Indels 13; Gaps 5;

QY 8 AFAAIVVSGSALAGV--VP---QWGGGNGHGG---GNSGPDSTLSIYQ---YGSANAA 56  
DB 10 AFAVVVSSCVADGVRTIPAGVQGSAGGLHNGGTAPSAANAANGSTTTAYDDRGTTGGQTAT 69  
QY 57 LALQSDARKSETTITQSGYNGADVGQGA 85  
DB 70 FVQOQGAQPEETTE--MGNAAEAAATGS 96

Search completed: August 2, 2004, 15:36:09  
Job time : 36.8 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: August 2, 2004, 14:48:33 ; Search time 167.9 Seconds

(without alignments)

877.809 Million cell updates/sec

Title: US-09-543-407-14

Perfect score: 775

Sequence: 1 MKLLKVAFAAIVVSGSALA.....HASVMVRQVGFNNATANYQ 151

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 6019581 seqs, 976053577 residues

Total number of hits satisfying chosen parameters: 6019581

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Pending Patents AA Main:\*

1: /cgn2\_6/ptodata/2/paa/PCTUS\_COMB.pep.\*

2: /cgn2\_6/ptodata/2/paa/US06\_COMB.pep.\*

3: /cgn2\_6/ptodata/2/paa/US07\_COMB.pep.\*

4: /cgn2\_6/ptodata/2/paa/US08\_COMB.pep.\*

5: /cgn2\_6/ptodata/2/paa/US081\_COMB.pep.\*

6: /cgn2\_6/ptodata/2/paa/US082\_COMB.pep.\*

7: /cgn2\_6/ptodata/2/paa/US083\_COMB.pep.\*

8: /cgn2\_6/ptodata/2/paa/US084\_COMB.pep.\*

9: /cgn2\_6/ptodata/2/paa/US085\_COMB.pep.\*

10: /cgn2\_6/ptodata/2/paa/US086\_COMB.pep.\*

11: /cgn2\_6/ptodata/2/paa/US087\_COMB.pep.\*

12: /cgn2\_6/ptodata/2/paa/US088\_COMB.pep.\*

13: /cgn2\_6/ptodata/2/paa/US089\_COMB.pep.\*

14: /cgn2\_6/ptodata/2/paa/US090\_COMB.pep.\*

15: /cgn2\_6/ptodata/2/paa/US091\_COMB.pep.\*

16: /cgn2\_6/ptodata/2/paa/US092\_COMB.pep.\*

17: /cgn2\_6/ptodata/2/paa/US093\_COMB.pep.\*

18: /cgn2\_6/ptodata/2/paa/US094\_COMB.pep.\*

19: /cgn2\_6/ptodata/2/paa/US095\_COMB.pep.\*

20: /cgn2\_6/ptodata/2/paa/US096\_COMB.pep.\*

21: /cgn2\_6/ptodata/2/paa/US097A\_COMB.pep.\*

22: /cgn2\_6/ptodata/2/paa/US097B\_COMB.pep.\*

23: /cgn2\_6/ptodata/2/paa/US098\_COMB.pep.\*

24: /cgn2\_6/ptodata/2/paa/US099A\_COMB.pep.\*

25: /cgn2\_6/ptodata/2/paa/US099B\_COMB.pep.\*

26: /cgn2\_6/ptodata/2/paa/US100\_COMB.pep.\*

27: /cgn2\_6/ptodata/2/paa/US101\_COMB.pep.\*

28: /cgn2\_6/ptodata/2/paa/US102\_COMB.pep.\*

29: /cgn2\_6/ptodata/2/paa/US103\_COMB.pep.\*

30: /cgn2\_6/ptodata/2/paa/US104\_COMB.pep.\*

31: /cgn2\_6/ptodata/2/paa/US106\_COMB.pep.\*

32: /cgn2\_6/ptodata/2/paa/US107\_COMB.pep.\*

33: /cgn2\_6/ptodata/2/paa/US60\_COMB.pep.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	ID	Description
------------	-------	-------	--------	----	-------------

1	775	100.0	151	19	US-09-543-407-14	Sequence 14, Appl
2	714	92.1	151	19	US-09-543-407-24	Sequence 24, Appl
3	696	89.8	151	19	US-09-543-407-12	Sequence 12, Appl
4	693	89.4	151	19	US-09-543-407-5	Sequence 5, Appl
5	692	89.3	151	6	US-08-233-642A-57	Sequence 57, Appl
6	659	85.0	151	19	US-09-543-407-30	Sequence 30, Appl
7	612	79.0	151	19	US-09-543-407-26	Sequence 26, Appl
8	611	78.8	151	19	US-09-543-407-18	Sequence 18, Appl
9	609	78.6	151	19	US-09-543-407-20	Sequence 20, Appl
10	606	78.2	131	19	US-09-543-407-31	Sequence 31, Appl
11	603	77.8	131	19	US-09-543-407-28	Sequence 28, Appl
12	602	77.7	151	19	US-09-543-407-22	Sequence 22, Appl
13	578	74.6	151	19	US-09-543-407-16	Sequence 16, Appl
14	528	68.1	151	19	US-09-543-407-7	Sequence 7, Appl
15	525	67.7	151	13	US-08-978-878-4	Sequence 4, Appl
16	525	67.7	151	21	US-09-741-873B-4	Sequence 2, Appl
17	523	67.5	151	33	US-60-352-946-2	Sequence 2, Appl
18	523	67.5	151	33	US-60-444-371-2	Sequence 55, Appl
19	514	66.3	120	6	US-08-233-642A-55	Sequence 34, Appl
20	474	61.2	109	19	US-09-543-407-34	Sequence 5834, Ap
21	472	60.9	158	16	US-09-252-691C-5834	Sequence 5834, Ap
22	472	60.9	158	16	US-09-252-691C-5834	Sequence 5834, Ap
23	472	60.9	158	30	US-10-417-886-5834	Sequence 2, Appl
24	447	57.7	131	13	US-08-978-878-2	Sequence 2, Appl
25	447	57.7	131	21	US-09-741-873B-2	Sequence 37, Appl
26	347	44.8	68	19	US-09-543-407-37	Sequence 35, Appl
27	343	44.3	109	19	US-09-543-407-35	Sequence 39, Appl
28	243.5	31.4	70	19	US-09-543-407-32	Sequence 32, Appl
29	237	30.6	48	19	US-09-543-407-39	Sequence 5833, Ap
30	115.5	14.9	186	16	US-09-252-691C-5833	Sequence 5833, Ap
31	115.5	14.9	186	16	US-09-252-691C-5833	Sequence 5833, Ap
32	115.5	14.9	186	30	US-10-417-886-5833	Sequence 8, Appl
33	104.5	13.5	151	19	US-09-543-407-8	Sequence 20638, A
34	104	13.4	445	29	US-10-369-493-20638	Sequence 20638, A
35	104	13.4	445	33	US-60-360-039-20638	Sequence 43013, A
36	98.5	12.7	151	19	US-09-543-407-6	Sequence 53421, A
37	98	12.6	210	28	US-10-219-999-43013	Sequence 53421, A
38	98	12.6	210	30	US-10-425-114-53421	Sequence 53421, A
39	98	12.6	210	30	US-10-425-114-53421	Sequence 53421, A
40	98	12.6	210	33	US-60-312-544-6765	Sequence 30520, A
41	98	12.6	210	33	US-60-324-109-30520	Sequence 245046, A
42	98	12.6	211	30	US-10-424-599-245046	Sequence 49412, A
43	95	12.3	186	28	US-10-282-122A-49412	Sequence 8, Appl
44	92	11.9	673	20	US-09-678-411-8	Sequence 10, Appl
45	92	11.9	949	20	US-09-678-411-10	

ALIGNMENTS

RESULT 1

US-09-543-407-14

/ Sequence 14, Application US/09543407

/ GENERAL INFORMATION:

/ APPLICANT: White, Aaron P.

/ APPLICANT: Doran, James L.

/ APPLICANT: Collinson, S. Karen

/ APPLICANT: Kay, William W.

/ TITLE OF INVENTION: BACTERIAL FIBRILLAR SYSTEM FOR PRESENTATION OF HETEROLOGOUS PEPTIDE SEQUENCES

/ FILE REFERENCE: 920043.406

/ CURRENT APPLICATION NUMBER: US/09/543.407

/ CURRENT FILING DATE: 2000-04-05

/ NUMBER OF SEQ ID NOS: 59

/ SOFTWARE: FastSeq for Windows Version 4.0

/ SEQ ID NO 14

/ LENGTH: 151

/ TYPE: PRT

/ ORGANISM: Artificial Sequence

/ FEATURE:

/ OTHER INFORMATION: Recombinant Salmonella enteritidis 3b afga

/ OTHER INFORMATION: sequence containing the replacement fragment

/ OTHER INFORMATION: encoding PT3 from Gp63 of Leishmania major.

US-09-543-407-14

Query Match 100.0%; Score 775; DB 19; Length 151;  
Best Local Similarity 100.0%; Pred. No. 6,1e-75;  
Matches 151; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MKLLKVAFAAIVVSGSALAGVVPQWGGGNGHNGGNSGPDSTLSIYQYGSANAALALQ 60  
DB 1 MKLLKVAFAAIVVSGSALAGVVPQWGGGNGHNGGNSGPDSTLSIYQYGSANAALALQ 60  
QY 61 SDARKSETTITQSGYNGADVGQADNSTIELTQNGFRNNATIDQWNAKNSDITVGYDQ 120  
DB 61 SDARKSETTITQSGYNGADVGQADNSTIELTQNGFRNNATIDQWNAKNSDITVGYDQ 120

QY 121 LVTRVVTHEMAHASVMVRQVGFNNATANQY 151  
DB 121 LVTRVVTHEMAHASVMVRQVGFNNATANQY 151

## RESULT 2

US-09-543-407-24

; Sequence 24, Application US/09543407  
; GENERAL INFORMATION:  
; APPLICANT: White, Aaron P.  
; APPLICANT: Doran, James L.  
; APPLICANT: Collinson, S. Karen  
; TITLE OF INVENTION: BACTERIAL FIMBRIAL SYSTEM FOR  
; FILE REFERENCE: 920043.406  
; CURRENT APPLICATION NUMBER: US/09/543,407  
; CURRENT FILING DATE: 2000-04-05  
; NUMBER OF SEQ ID NOS: 59  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 24  
; LENGTH: 151  
; TYPE: PRT  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: Recombinant Salmonella enteritidis 3b afgA  
; OTHER INFORMATION: sequence containing the replacement fragment  
; OTHER INFORMATION: encoding PT3 from GP63 of Leishmania major.  
US-09-543-407-24

Query Match 92.1%; Score 714; DB 19; Length 151;  
Best Local Similarity 91.1%; Pred. No. 2.4e-68;  
Matches 144; Conservative 0; Mismatches 0; Indels 14; Gaps 2;

QY 1 MKLLKVAFAAIVVSGSALAGVVPQWGGGNGHNGGNSGPDSTLSIYQYGSANAALALQ 60  
DB 1 MKLLKVAFAAIVVSGSALAGVVPQWGGGNGHNGGNSGPDSTLSIYQYGSANAALALQ 60  
QY 61 SDARKSETTITQSGYNGADVGQADNSTIELTQNGFRNNATIDQWNAKNSDITVGYDQ 120  
DB 61 SDARKSETTITQSGYNGADVGQADNSTIELTQNGFRNNATIDQWNAKNSDITVGYDQ 113

QY 121 LVTRVVTHEMAHA-----SVNVRQVGFNNATANQY 151  
DB 114 LVTRVVTHEMAHANQTASDSSVMVRQVGFNNATANQY 151

## RESULT 3

US-09-543-407-12

; Sequence 12, Application US/09543407  
; GENERAL INFORMATION:  
; APPLICANT: White, Aaron P.  
; APPLICANT: Doran, James L.  
; APPLICANT: Collinson, S. Karen  
; TITLE OF INVENTION: BACTERIAL FIMBRIAL SYSTEM FOR  
; FILE REFERENCE: 920043.406  
; CURRENT APPLICATION NUMBER: US/09/543,407

CURRENT FILING DATE: 2000-04-05

; NUMBER OF SEQ ID NOS: 59  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 12

; LENGTH: 151  
; TYPE: PRT  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: Recombinant Salmonella enteritidis 3b afgA  
; OTHER INFORMATION: sequence containing the replacement fragment  
; OTHER INFORMATION: encoding PT3 from GP63 of Leishmania major.  
US-09-543-407-12

Query Match 89.8%; Score 696; DB 19; Length 151;  
Best Local Similarity 87.6%; Pred. No. 2.1e-66;  
Matches 141; Conservative 0; Mismatches 0; Indels 20; Gaps 2;

QY 1 MKLLKVAFAAIVVSGSALAGVVPQWGGGNGHNGGNSGPDSTLSIYQYGSANAALALQ 60  
DB 1 MKLLKVAFAAIVVSGSALAGVVPQWGGGNGHNGGNSGPDSTLSIYQYGSANAALALQ 60

QY 61 SDARKSETTITQSGYNGADVGQADNSTIELTQNGFRNNATIDQWNAKNSDITVGYDQ 117  
DB 61 SDARKSETTITQSGYNGADVGQADNSTIELTQNGFRNNATIDQWNAKNSDITVGYDQ 120

QY 118 -----YDOLVTRVVTHEMAHASVMVRQVGFNNATANQY 151  
DB 121 NNAALVNYDQLVTRVVTHEMAHA-----NNATANQY 151

## RESULT 4

US-09-543-407-5

; Sequence 5, Application US/09543407  
; GENERAL INFORMATION:  
; APPLICANT: White, Aaron P.  
; APPLICANT: Doran, James L.  
; APPLICANT: Collinson, S. Karen  
; APPLICANT: Kay, William W.  
; TITLE OF INVENTION: BACTERIAL FIMBRIAL SYSTEM FOR  
; FILE REFERENCE: 920043.406  
; CURRENT APPLICATION NUMBER: US/09/543,407  
; CURRENT FILING DATE: 2000-04-05  
; NUMBER OF SEQ ID NOS: 59  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 5  
; LENGTH: 151  
; TYPE: PRT  
; ORGANISM: Salmonella enteritidis  
US-09-543-407-5

Query Match 89.4%; Score 693; DB 19; Length 151;  
Best Local Similarity 90.7%; Pred. No. 4.4e-66;  
Matches 137; Conservative 3; Mismatches 11; Indels 0; Gaps 0;

QY 1 MKLLKVAFAAIVVSGSALAGVVPQWGGGNGHNGGNSGPDSTLSIYQYGSANAALALQ 60  
DB 1 MKLLKVAFAAIVVSGSALAGVVPQWGGGNGHNGGNSGPDSTLSIYQYGSANAALALQ 60  
QY 61 SDARKSETTITQSGYNGADVGQADNSTIELTQNGFRNNATIDQWNAKNSDITVGYDQ 120  
DB 61 SDARKSETTITQSGYNGADVGQADNSTIELTQNGFRNNATIDQWNAKNSDITVGYDQ 120

QY 121 LVTRVVTHEMAHASVMVRQVGFNNATANQY 151  
DB 121 NNAALVNYDQLVTRVVTHEMAHA-----NNATANQY 151

## RESULT 5

US-08-233-642A-57

; Sequence 57, Application US/08233642A  
; GENERAL INFORMATION:  
; APPLICANT: Kay, William W.

APPLICANT: Collinson, S. Karen  
 APPLICANT: Clouthier, Sharon C.  
 APPLICANT: Doran, James L.  
 TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR SALMONELLA-  
 BASED VACCINES  
 NUMBER OF SEQUENCES: 58  
 NUMBER OF SEQUENCES: 58  
 CORRESPONDENCE ADDRESS:  
 ADDRESSEE: Seed and Berry  
 STREET: 6300 Columbia Center, 701 Fifth Avenue  
 CITY: Seattle  
 STATE: Washington  
 COUNTRY: U.S.A.  
 ZIP: 98104-7092  
 COMPUTER READABLE FORM:  
 MEDIUM TYPE: Floppy disk  
 COMPUTER: IBM PC compatible  
 OPERATING SYSTEM: PC-DOS/MS-DOS  
 SOFTWARE: Patent In Release #1.0, Version #1.25  
 CURRENT APPLICATION DATA:  
 APPLICATION NUMBER: US/08/233,642A  
 FILING DATE: 26-APR-1994  
 CLASSIFICATION: 424  
 ATTORNEY/AGENT INFORMATION:  
 NAME: King, Joshua  
 REGISTRATION NUMBER: 35,570  
 REFERENCE/DOCKET NUMBER: 920043.403C3  
 TELECOMMUNICATION INFORMATION:  
 TELEPHONE: (206) 622-4900  
 TELEFAX: (206) 682-6031  
 TELEX: 3723836 SEEDANBERRY  
 INFORMATION FOR SEQ ID NO: 57:  
 SEQUENCE CHARACTERISTICS:  
 LENGTH: 151 amino acids  
 TYPE: amino acid  
 TOPOLOGY: linear  
 MOLECULE TYPE: protein  
 US-08-233-642A-57

Query Match 89.3%; Score 692; DB 6; Length 151;  
 Best Local Similarity 90.7%; Pred. No. 5.6e-66;  
 Matches 137; Conservative 3; Mismatches 11; Indels 0; Gaps 0;  
 QY 1 MLLKVAFAAIVVSGSALAGVVPQWGGGNGHNGGNSGPDSTLSIYQYGSANAALALQ 60  
 Db 1 MLLKVAFAAIVVSGSALAGVVPQWGGGNGHNGGNSGPDSTLSIYQYGSANAALALQ 60  
 QY 61 SDARKSETTITQSGYNGADVGQADNSTIELTQNGFRNNATIDQWNAKNSDITVGYDQ 120  
 Db 61 SDARKSETTITQSGYNGADVGQADNSTIELTQNGFRNNATIDQWNAKNSDITVGYDQ 120  
 QY 121 LVTRVVTHEMAHSAVMVRQVGFNNATANQY 151  
 Db 121 NNPAVNQTASDSSVMVRQVGFNNATANQY 151

RESULT 6  
 US-09-543-407-30  
 ; Sequence 30, Application US/09543407  
 ; GENERAL INFORMATION:  
 ; APPLICANT: White, Aaron P.  
 ; APPLICANT: Doran, James L.  
 ; APPLICANT: Collinson, S. Karen  
 ; APPLICANT: Kay, William W.  
 ; TITLE OF INVENTION: BACTERIAL FIMBRIAL SYSTEM FOR  
 ; PRESENTATION OF HETEROLOGOUS PEPTIDE SEQUENCES  
 ; FILE REFERENCE: 920043.406  
 ; CURRENT APPLICATION NUMBER: US/09/543,407  
 ; CURRENT FILING DATE: 2000-04-05  
 ; NUMBER OF SEQ ID NOS: 59  
 ; SOFTWARE: FastSeq for Windows Version 4.0  
 ; SEQ ID NO 30  
 ; LENGTH: 151

; TYPE: PRT  
 ; ORGANISM: Artificial Sequence  
 ; FEATURE:  
 ; OTHER INFORMATION: Recombinant Salmonella enteritidis 3b afga  
 ; OTHER INFORMATION: sequence containing the replacement fragment  
 ; OTHER INFORMATION: encoding PT3 from GP63 of Leishmania major.  
 US-09-543-407-30

Query Match 85.0%; Score 659; DB 19; Length 151;  
 Best Local Similarity 81.9%; Pred. No. 2.1e-62;  
 Matches 136; Conservative 0; Mismatches 0; Indels 30; Gaps 2;  
 QY 1 MLLKVAFAAIVVSGSALAGVVPQWGGGNGHNGGNSGPDSTLSIYQYGSANAALALQ 60  
 Db 1 MLLKVAFAAIVVSGSALAGVVPQWGGGNGHNGGNSGPDSTLSIYQYGSANAALALQ 60  
 QY 61 SDARKSETTITQSGYNGADVGQADNSTIELTQNGFRNNATIDQWNAKNSDITVGYDQ 120  
 Db 61 SDARKSETTITQSGYNGADVGQADNSTIELTQNGFRNNATIDQWNAKNSDITVGYDQ 120  
 QY 121 LVTRVVTHEMAHA-----SYMVRQVGFNNATANQY 151  
 Db 106 LVTRVVTHEMAHAGNNAALVNQTASDSSVMVRQVGFNNATANQY 151

RESULT 7  
 US-09-543-407-26  
 ; Sequence 26, Application US/09543407  
 ; GENERAL INFORMATION:  
 ; APPLICANT: White, Aaron P.  
 ; APPLICANT: Doran, James L.  
 ; APPLICANT: Collinson, S. Karen  
 ; APPLICANT: Kay, William W.  
 ; TITLE OF INVENTION: BACTERIAL FIMBRIAL SYSTEM FOR  
 ; PRESENTATION OF HETEROLOGOUS PEPTIDE SEQUENCES  
 ; FILE REFERENCE: 920043.406  
 ; CURRENT APPLICATION NUMBER: US/09/543,407  
 ; CURRENT FILING DATE: 2000-04-05  
 ; SOFTWARE: FastSeq for Windows Version 4.0  
 ; SEQ ID NO 26  
 ; LENGTH: 151  
 ; TYPE: PRT  
 ; ORGANISM: Artificial Sequence  
 ; FEATURE:  
 ; OTHER INFORMATION: Recombinant Salmonella enteritidis 3b afga  
 ; OTHER INFORMATION: sequence containing the replacement fragment  
 ; OTHER INFORMATION: encoding PT3 from GP63 of Leishmania major.  
 US-09-543-407-26

Query Match 79.0%; Score 612; DB 19; Length 151;  
 Best Local Similarity 81.5%; Pred. No. 2.5e-57;  
 Matches 123; Conservative 5; Mismatches 23; Indels 0; Gaps 0;  
 QY 1 MLLKVAFAAIVVSGSALAGVVPQWGGGNGHNGGNSGPDSTLSIYQYGSANAALALQ 60  
 Db 1 MLLKVAFAAIVVSGSALAGVVPQWGGGNGHNGGNSGPDSTLSIYQYGSANAALALQ 60  
 QY 61 SDARKSETTITQSGYNGADVGQADNSTIELTQNGFRNNATIDQWNAKNSDITVGYDQ 120  
 Db 61 LVTRVVTHEMAHAGYNGADVGQADNSTIELTQNGFRNNATIDQWNAKNSDITVGYDQ 120  
 QY 121 LVTRVVTHEMAHASVMVRQVGFNNATANQY 151  
 Db 121 NNPAVNQTASDSSVMVRQVGFNNATANQY 151

RESULT 8  
 US-09-543-407-18  
 ; Sequence 18, Application US/09543407  
 ; GENERAL INFORMATION:  
 ; APPLICANT: White, Aaron P.  
 ; APPLICANT: Doran, James L.

APPLICANT: Collinson, S. Karen  
APPLICANT: Kay, William W.  
TITLE OF INVENTION: BACTERIAL FIMBRIAL SYSTEM FOR  
FILE REFERENCE: 920043.406  
CURRENT FILING DATE: 2000-04-05  
NUMBER OF SEQ ID NOS: 59  
SOFTWARE: FastSeq for Windows Version 4.0  
SEQ ID NO 18  
LENGTH: 151  
TYPE: PRT  
ORGANISM: Artificial Sequence  
FEATURE:  
OTHER INFORMATION: Recombinant Salmonella enteritidis 3b afgA  
OTHER INFORMATION: sequence containing the replacement fragment  
OTHER INFORMATION: encoding PT3 from Gp63 of Leishmania major.  
US-09-543-407-18

Query Match 78.8%; Score 611; DB 19; Length 151;  
Best Local Similarity 81.5%; Pred. No. 3.2e-57;  
Matches 123; Conservative 6; Mismatches 22; Indels 0; Gaps 0;

Qy 1 MKLLKVAFAAIVVSGSALAGVVPQGGGNGHNGGSGPDSITLSIYQYGSANAALALQ 60  
Db 1 MKLLKVAFAAIVVSGSALAGVVPQGGGNGHNGGSGPDSITLSIYQYGSANAALALQ 60  
Qy 61 SDARKSETTITQSGYNGADVGQADNSTIETQNGFRNNATIDQWNAKNSDITVQYDQ 120  
Db 61 SDARKSETTITQSGYNGADVGQADNSTIETQNGFRNNATIDQWNAKNSDITVQYGG 120  
Qy 121 LVTRVVTHEMAHASVMVRQVGFNNATANQY 151  
Db 121 NNAALVNQTASDSSVMVRQVGFNNATANQY 151

RESULT 9  
US-09-543-407-20  
Sequence 20, Application US/09543407  
GENERAL INFORMATION:  
APPLICANT: White, Aaron P.  
APPLICANT: Doran, James L.  
APPLICANT: Collinson, S. Karen  
APPLICANT: Kay, William W.  
TITLE OF INVENTION: BACTERIAL FIMBRIAL SYSTEM FOR  
FILE REFERENCE: 920043.406  
CURRENT FILING DATE: 2000-04-05  
NUMBER OF SEQ ID NOS: 59  
SOFTWARE: FastSeq for Windows Version 4.0  
SEQ ID NO 20  
LENGTH: 151  
TYPE: PRT  
ORGANISM: Artificial Sequence  
FEATURE:  
OTHER INFORMATION: Recombinant Salmonella enteritidis 3b afgA  
OTHER INFORMATION: sequence containing the replacement fragment  
OTHER INFORMATION: encoding PT3 from Gp63 of Leishmania major.  
US-09-543-407-20

Query Match 78.6%; Score 609; DB 19; Length 151;  
Best Local Similarity 81.5%; Pred. No. 5.2e-57;  
Matches 123; Conservative 6; Mismatches 22; Indels 0; Gaps 0;

Qy 1 MKLLKVAFAAIVVSGSALAGVVPQGGGNGHNGGSGPDSITLSIYQYGSANAALALQ 60  
Db 1 MKLLKVAFAAIVVSGSALAGVVPQGGGNGHNGGSGPDSITLSIYQYGSANAALALQ 60  
Qy 61 SDARKSETTITQSGYNGADVGQADNSTIETQNGFRNNATIDQWNAKNSDITVQYDQ 120  
Db 61 SDARKYDQLVTRVVTHEMAHAGQADNSTIETQNGFRNNATIDQWNAKNSDITVQYGG 120

Qy 121 LVTRVVTHEMAHASVMVRQVGFNNATANQY 151  
Db 121 NNAALVNQTASDSSVMVRQVGFNNATANQY 151

RESULT 10  
US-09-543-407-31  
Sequence 31, Application US/09543407  
GENERAL INFORMATION:  
APPLICANT: White, Aaron P.  
APPLICANT: Doran, James L.  
APPLICANT: Collinson, S. Karen  
APPLICANT: Kay, William W.  
TITLE OF INVENTION: BACTERIAL FIMBRIAL SYSTEM FOR  
FILE REFERENCE: 920043.406  
CURRENT FILING DATE: 2000-04-05  
NUMBER OF SEQ ID NOS: 59  
SOFTWARE: FastSeq for Windows Version 4.0  
SEQ ID NO 31  
LENGTH: 131  
TYPE: PRT  
ORGANISM: Salmonella enteritidis  
US-09-543-407-31

Query Match 78.2%; Score 606; DB 19; Length 131;  
Best Local Similarity 89.3%; Pred. No. 9e-57;  
Matches 117; Conservative 3; Mismatches 11; Indels 0; Gaps 0;

Qy 21 GVVPQGGGNGHNGGSGPDSITLSIYQYGSANAALALQSDARKSETTITQSGYNGAD 80  
Db 1 GVVPQGGGNGHNGGSGPDSITLSIYQYGSANAALALQSDARKSETTITQSGYNGAD 60  
Qy 81 VGOADNSTIETQNGFRNNATIDQWNAKNSDITVQYDQVTRVVTHEMAHASVMVRQV 140  
Db 61 VGOADNSTIETQNGFRNNATIDQWNAKNSDITVQYDQVTRVVTHEMAHASVMVRQV 120  
Qy 141 GFNNATANQY 151  
Db 121 GFNNATANQY 131

RESULT 11  
US-09-543-407-28  
Sequence 28, Application US/09543407  
GENERAL INFORMATION:  
APPLICANT: White, Aaron P.  
APPLICANT: Doran, James L.  
APPLICANT: Collinson, S. Karen  
APPLICANT: Kay, William W.  
TITLE OF INVENTION: BACTERIAL FIMBRIAL SYSTEM FOR  
FILE REFERENCE: 920043.406  
CURRENT FILING DATE: 2000-04-05  
NUMBER OF SEQ ID NOS: 59  
SOFTWARE: FastSeq for Windows Version 4.0  
SEQ ID NO 28  
LENGTH: 151  
TYPE: PRT  
ORGANISM: Artificial Sequence  
FEATURE:  
OTHER INFORMATION: Recombinant Salmonella enteritidis 3b afgA  
OTHER INFORMATION: sequence containing the replacement fragment  
OTHER INFORMATION: encoding PT3 from Gp63 of Leishmania major.  
US-09-543-407-28

Query Match 77.8%; Score 603; DB 19; Length 151;  
Best Local Similarity 80.8%; Pred. No. 2.3e-56;  
Matches 122; Conservative 4; Mismatches 25; Indels 0; Gaps 0;

Qy 1 MKLLKVAFAAIVVSGSALAGVVPQGGGNGHNGGSGPDSITLSIYQYGSANAALALQ 60

```
Db 1 MKLLKVAFAAIVVSGSALAGVVPQGGGNGHNGGNSGPDSTLSIYQYGSANAALALQ 60
QY 61 SDARKSETTITQSGYNGADVGGADNSTIELTQNGFRNNATIDQWNAKNSDITVGYDQ 120
Db 61 SDARKSETTITQSGYNGADYDQLVTRVVTHEMAHAFRNATIDQWNAKNSDITVGYGG 120
QY 121 LVTRVVTHEMAHASVMVRQVGFNNATANQY 151
Db 121 NNAALVNQTASDSSVMVRQVGFNNATANQY 151

RESULT 12
US-09-543-407-22
; Sequence 22, Application US/09543407
; GENERAL INFORMATION:
; APPLICANT: White, Aaron P.
; APPLICANT: Doran, James L.
; APPLICANT: Collinson, S. Karen
; APPLICANT: Kay, William W.
; TITLE OF INVENTION: BACTERIAL FIMBRIAL SYSTEM FOR
; TITLE OF INVENTION: PRESENTATION OF HETEROLOGOUS PEPTIDE SEQUENCES
; FILE REFERENCE: 920043.406
; CURRENT APPLICATION NUMBER: US/09/543,407
; CURRENT FILING DATE: 2000-04-05
; NUMBER OF SEQ ID NOS: 59
; SOFTWARE: Fast-SEQ for Windows Version 4.0
; SEQ ID NO 22
; LENGTH: 151
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Recombinant Salmonella enteritidis 3b afgA
; OTHER INFORMATION: sequence containing the replacement fragment
; OTHER INFORMATION: encoding PT3 from GP63 of Leishmania major.
US-09-543-407-22

Query Match 77.7%; Score 602; DB 19; Length 151;
Best Local Similarity 81.5%; Pred. No. 3e-56;
Matches 123; Conservative 5; Mismatches 23; Indels 0; Gaps 0;

QY 1 MKLLKVAFAAIVVSGSALAGVVPQGGGNGHNGGNSGPDSTLSIYQYGSANAALALQ 60
Db 1 MKLLKVAFAAIVVSGSALAGVVPQGGGNGHNGGNSGPDSTLSIYQYGSANAALALQ 60
QY 61 SDARKSETTITQSGYNGADVGGADNSTIELTQNGFRNNATIDQWNAKNSDITVGYDQ 120
Db 61 SDARKSETTITQSGYNGADVGGADNYDQLVTRVVTHEMAHADQWNAKNSDITVGYGG 120
QY 121 LVTRVVTHEMAHASVMVRQVGFNNATANQY 151
Db 121 NNAALVNQTASDSSVMVRQVGFNNATANQY 151

RESULT 13
US-09-543-407-16
; Sequence 16, Application US/09543407
; GENERAL INFORMATION:
; APPLICANT: White, Aaron P.
; APPLICANT: Doran, James L.
; APPLICANT: Collinson, S. Karen
; APPLICANT: Kay, William W.
; TITLE OF INVENTION: BACTERIAL FIMBRIAL SYSTEM FOR
; TITLE OF INVENTION: PRESENTATION OF HETEROLOGOUS PEPTIDE SEQUENCES
; FILE REFERENCE: 920043.406
; CURRENT APPLICATION NUMBER: US/09/543,407
; CURRENT FILING DATE: 2000-04-05
; NUMBER OF SEQ ID NOS: 59
; SOFTWARE: Fast-SEQ for Windows Version 4.0
; SEQ ID NO 16
; LENGTH: 151
; TYPE: PRT
; ORGANISM: Artificial Sequence
```

```
; FEATURE:
; OTHER INFORMATION: Recombinant Salmonella enteritidis 3b afgA
; OTHER INFORMATION: sequence containing the replacement fragment
; OTHER INFORMATION: encoding PT3 from GP63 of Leishmania major.
US-09-543-407-16

Query Match 74.6%; Score 578; DB 19; Length 151;
Best Local Similarity 80.8%; Pred. No. 1.2e-53;
Matches 122; Conservative 6; Mismatches 23; Indels 0; Gaps 0;

QY 1 MKLLKVAFAAIVVSGSALAGVVPQGGGNGHNGGNSGPDSTLSIYQYGSANAALALQ 60
Db 1 MKLLKVAFAAIVVSGSALAGVYDQLVTRVVTHEMAHASGPDSTLSIYQYGSANAALALQ 60
QY 61 SDARKSETTITQSGYNGADVGGADNSTIELTQNGFRNNATIDQWNAKNSDITVGYDQ 120
Db 61 SDARKSETTITQSGYNGADVGGADNSTIELTQNGFRNNATIDQWNAKNSDITVGYGG 120
QY 121 LVTRVVTHEMAHASVMVRQVGFNNATANQY 151
Db 121 NNAALVNQTASDSSVMVRQVGFNNATANQY 151

RESULT 14
US-09-543-407-7
; Sequence 7, Application US/09543407
; GENERAL INFORMATION:
; APPLICANT: White, Aaron P.
; APPLICANT: Doran, James L.
; APPLICANT: Collinson, S. Karen
; APPLICANT: Kay, William W.
; TITLE OF INVENTION: BACTERIAL FIMBRIAL SYSTEM FOR
; TITLE OF INVENTION: PRESENTATION OF HETEROLOGOUS PEPTIDE SEQUENCES
; FILE REFERENCE: 920043.406
; CURRENT APPLICATION NUMBER: US/09/543,407
; CURRENT FILING DATE: 2000-04-05
; NUMBER OF SEQ ID NOS: 59
; SOFTWARE: Fast-SEQ for Windows Version 4.0
; SEQ ID NO 7
; LENGTH: 151
; TYPE: PRT
; ORGANISM: Escherichia coli
US-09-543-407-7

Query Match 68.1%; Score 528; DB 19; Length 151;
Best Local Similarity 68.9%; Pred. No. 2.9e-48;
Matches 104; Conservative 20; Mismatches 27; Indels 0; Gaps 0;

QY 1 MKLLKVAFAAIVVSGSALAGVVPQGGGNGHNGGNSGPDSTLSIYQYGSANAALALQ 60
Db 1 MKLLKVAFAAIVVSGSALAGVVPQGGGNGHNGGNSGNSLNLYQYGGNSALALQ 60
QY 61 SDARKSETTITQSGYNGADVGGADNSTIELTQNGFRNNATIDQWNAKNSDITVGYDQ 120
Db 61 TDARNSDLTITQHGCGGADVGGSDSSIDLTLQRFNSATLDQWNGKNSMTVKQFGG 120
QY 121 LVTRVVTHEMAHASVMVRQVGFNNATANQY 151
Db 121 GNGAAVDQTASNSVNVTVQVGFNNATAHQY 151

RESULT 15
US-08-978-878-4
; Sequence 4, Application US/08978878
; GENERAL INFORMATION:
; APPLICANT: NORMARK, Staffan
; APPLICANT: OLSEN, Arne
; TITLE OF INVENTION: FIBRONECTIN BINDING PROTEIN AS WELL AS ITS PREPARATION
; FILE REFERENCE: 012889-081
; CURRENT APPLICATION NUMBER: US/08/978,878
; CURRENT FILING DATE: 1997-11-26
; EARLIER APPLICATION NUMBER: SE 8601723-1
; EARLIER FILING DATE: 1988-05-06
```

```
; EARLIER APPLICATION NUMBER: US 07/347,189
; EARLIER FILING DATE: 1989-05-04
; EARLIER APPLICATION NUMBER: US 07/789,437
; EARLIER FILING DATE: 1991-11-06
; EARLIER APPLICATION NUMBER: US 07/970,846
; EARLIER FILING DATE: 1992-11-03
; EARLIER APPLICATION NUMBER: US 08/187,865
; EARLIER FILING DATE: 1994-01-28
; EARLIER APPLICATION NUMBER: US 08/318,519
; EARLIER FILING DATE: 1994-10-05
; EARLIER APPLICATION NUMBER: US 08/495,959
; EARLIER FILING DATE: 1995-06-28
; NUMBER OF SEQ ID NOS: 10
; SOFTWARE: Patentin Ver. 2.0
; SEQ ID NO 4
; LENGTH: 151
; TYPE: PRT
; ORGANISM: Escherichia coli
US-08-978-878-4
```

```
Query Match          67.7%; Score 525; DB 13; Length 151;
Best Local Similarity 68.2%; Pred. No. 6.1e-48;
Matches 103; Conservative 21; Mismatches 27; Indels 0; Gaps 0;

QY      1  MKLLKVAFAAIVVSGSALAGVVPQWGGGNNHGGNGSSGPDSTLSIYYGSANAALALQ 60
      |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db      1  MKLLKVAATAAIVFSGSAGVAVVPPQYGGGNGHGGGNGSGPNSLNLYQYGGGNSALALQ 60

QY      61  SDARKSETTITQSGYNGADVGQADNSTIELTONGFRNNATIDOWNAKNSDITYGYDQ 120
      :|||:  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db      61  TDARNSDLTITQHGNGGADVGQGGSDSSIDLTLQRFNGSATLDDQWNGKNSNTVQKQGG 120

QY      121  LVTRVVTHEMAHASVMVROVQFGNNATANCY 151
      |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db      121  GNGAAYDQFASNSSVNVTVQFGNNATAHCY 151
```

```
Search completed: August 2, 2004, 15:26:42
Job time : 167.9 secs
```



GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: August 2, 2004, 14:49:38 ; Search time 17.8 Seconds  
(without alignments)  
888.146 Million cell updates/sec

Title: US-09-543-407-14  
Perfect score: 775  
Sequence: 1 MKLLKVAAPAAIVSGSALA.....HASVMVRQVGFNNATANQY 151

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 601315 seqs, 104695340 residues

Total number of hits satisfying chosen parameters: 601315

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Pending Patents AA New.\*  
1: /cgn2\_6/ptodata/2/paa/PCT NEW COMB.pcp.\*  
2: /cgn2\_6/ptodata/2/paa/US05 NEW COMB.pcp.\*  
3: /cgn2\_6/ptodata/2/paa/US07 NEW COMB.pcp.\*  
4: /cgn2\_6/ptodata/2/paa/US08 NEW COMB.pcp.\*  
5: /cgn2\_6/ptodata/2/paa/US09 NEW COMB.pcp.\*  
6: /cgn2\_6/ptodata/2/paa/US10 NEW COMB.pcp.\*  
7: /cgn2\_6/ptodata/2/paa/US60 NEW COMB.pcp.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	525	67.7	151	5	US-09-741-873C-4
2	447	57.7	131	5	US-09-741-873C-2
3	92	11.9	299	6	US-10-170-205E-35751
4	92	11.9	1203	6	US-10-170-205E-741
5	92	11.9	1327	1	PCT-US04-02338-49
6	90.5	11.7	258	6	US-10-425-115-300390
7	89.5	11.5	258	6	US-10-425-115-301334
8	89	11.5	258	6	US-10-425-115-312468
9	88	11.4	520	6	US-10-479-638-21
10	86.5	11.2	527	6	US-10-468-356-348
11	86	11.1	179	6	US-10-425-115-346132
12	85.5	11.0	1048	1	PCT-US04-05654-2086
13	82.5	10.6	535	1	PCT-US04-21492-88
14	82	10.6	758	7	US-60-565-632-9194
15	82	10.6	758	7	US-60-579-062-9194
16	82	10.6	1095	1	PCT-US03-24982A-273
17	82	10.6	2228	6	US-10-745-237-96
18	81.5	10.5	376	6	US-10-491-733-2
19	81.5	10.5	508	6	US-10-425-115-285216
20	81.5	10.5	532	6	US-10-425-115-285214
21	81	10.5	131	7	US-60-565-632-11109
22	81	10.5	131	7	US-60-579-062-11109
23	81	10.5	201	6	US-10-425-115-309662
24	81	10.5	376	6	US-10-425-115-342526
25	80.5	10.4	246	6	US-10-854-439-511
26	80.5	10.4	388	5	US-09-248-796A-17306

27 80.5 10.4 873 5 US-09-952-267B-13 Sequence 13, Appl  
28 80.5 10.4 873 6 US-10-872-768-13 Sequence 13, Appl  
29 80.5 10.4 873 6 US-10-872-769-13 Sequence 13, Appl  
30 80 10.3 892 5 US-09-952-267B-5 Sequence 5, Appl  
31 80 10.3 892 6 US-10-872-768-5 Sequence 5, Appl  
32 80 10.3 892 6 US-10-872-769-5 Sequence 5, Appl  
33 79 10.2 132 6 US-10-425-115-351875 Sequence 351875,  
34 79 10.2 586 1 PCT-US03-24982A-317 Sequence 317, App  
35 78.5 10.1 443 6 US-10-100-683-7608 Sequence 7608, Ap  
36 78.5 10.1 573 7 US-60-565-632-7907 Sequence 7907, Ap  
37 78.5 10.1 573 7 US-60-579-062-7907 Sequence 7907, Ap  
38 78.5 10.1 659 1 PCT-US04-09385-4 Sequence 4, Appl  
39 78.5 10.1 659 6 US-10-809-790-4 Sequence 4, Appl  
40 78.5 10.1 839 1 PCT-US04-12070-2 Sequence 2, Appl  
41 78.5 10.1 998 1 PCT-US04-21492-198 Sequence 198, App  
42 78 10.1 147 6 US-10-425-115-193207 Sequence 193207,  
43 77.5 10.0 196 6 US-10-425-115-254355 Sequence 254355,  
44 77.5 10.0 207 6 US-10-425-115-329691 Sequence 329691,  
45 77.5 10.0 269 6 US-10-425-115-190187 Sequence 190187,

## ALIGNMENTS

## RESULT 1

US-09-741-873C-4  
; Sequence 4, Application US/09741873C  
; GENERAL INFORMATION:  
; APPLICANT: Normark, Staffan  
; TITLE OF INVENTION: Fibronectin Binding Protein As Well As Its Preparation  
; FILE REFERENCE: 012889-084  
; CURRENT APPLICATION NUMBER: US/09/741,873C  
; CURRENT FILING DATE: 2000-12-22  
; PRIOR APPLICATION NUMBER: SE 8801723-1  
; PRIOR FILING DATE: 1998-05-06  
; PRIOR APPLICATION NUMBER: US 08/978,878  
; PRIOR FILING DATE: 1997-11-26  
; PRIOR APPLICATION NUMBER: US 07/347,189  
; PRIOR FILING DATE: 1989-05-04  
; PRIOR APPLICATION NUMBER: US 07/789,437  
; PRIOR FILING DATE: 1991-11-06  
; PRIOR APPLICATION NUMBER: US 07/970,846  
; PRIOR FILING DATE: 1992-11-03  
; PRIOR APPLICATION NUMBER: US 08/187,865  
; PRIOR FILING DATE: 1994-01-28  
; PRIOR APPLICATION NUMBER: US 08/318,519  
; NUMBER OF SEQ ID NOS: 11  
; SOFTWARE: Patent version 3.0  
; SEQ ID NO 4  
; LENGTH: 151  
; TYPE: PRT  
; ORGANISM: Escherichia coli  
US-09-741-873C-4

Query Match 67.7%; Score 525; DB 5; Length 151;

Best Local Similarity 68.2%; Pred. No. 2.3e-38;

Matches 103; Conservative 21; Mismatches 27; Indels 0; Gaps 0;

QY 1 MKLLKVAAPAAIVSGSALAAGVVPQGGGHNHGGNSGPDSTLSIYQYGSANAALAQ 60  
DB 1 MKLLKVAAPAAIVSGSALAAGVVPQGGGHNHGGNSGPDSTLSIYQYGSANAALAQ 60  
QY 61 SPARKSETTITQSGYNGADVGQGDADNSTIELTQNGFRNNATIDOWNAKNSDITVQYDQ 120  
DB 61 TDARNSLDITITQGGGNGADVGQGDSSIDILTORFGNSATLTDWNGKNSMTVKQFGG 120  
QY 121 LVTRVVTHEMAHASYVWVQVGFNNATANQY 151  
DB 121 GNGAAVDQTASNSSVNVTVQVGFNNATANQY 151



```
QY      15 SCSALAGVVPWGGCGNHGCGSSGPDSTLSIYQGSAAALALQSDARKSET-TITQS 73
       : : | | : | : | : | : | : | : | : | : | : | : | : | : | : | :
rb     347 AGCGAAAAAASACACAGAGVGCGVAGGSS-----ISYGATSSSATSSASSRSRGIVTSG 402
```



APPLICANT: Diversa Corporation  
APPLICANT: Steer, Brian  
APPLICANT: Callen, Walter  
APPLICANT: Healey, Shaun  
APPLICANT: Pulliam, Derrick  
TITLE OF INVENTION: GLUCANASES, NUCLEIC ACIDS ENCODING THEM AND METHODS FOR MAKING AN  
TITLE OF INVENTION: THEM  
FILE REFERENCE: 564462009540  
CURRENT APPLICATION NUMBER: PCT/US04/21492  
CURRENT FILING DATE: 2004-07-20  
PRIOR APPLICATION NUMBER: 60/484,725  
PRIOR FILING DATE: 2003-07-02  
NUMBER OF SEQ ID NOS: 518  
SOFTWARE: Patent in version 3.1  
SEQ ID NO 88  
LENGTH: 535  
TYPE: PRT  
ORGANISM: Unknown  
FEATURE:  
OTHER INFORMATION: Obtained from an environmental sample.  
NAME/KEY: SIGNAL  
LOCATION: (1)...(27)  
PCT-US04-21492-88

Query Match 10.6%; Score 82.5; DB 1; Length 535;  
Best Local Similarity 26.8%; Pred. No. 29;  
Matches 38; Conservative 18; Mismatches 58; Indels 31; Gaps 7;  
QY 12 IVVSG-----SALAGVVP--QWGGGN-----HNGG-----GNSGGPDSTLSIYQYGA 53  
DB 361 IVVNGTQRAENGSVNTGWANNQCGSGNSELHCGNVIISFGVSGSSSSSSSSSSSS 420  
QY 54 NAALQSDARKSETTITQ-----SGVNGADVGGADNSTIELT-QNGFRRNATIQQ 105  
DB 421 SSSSTSS 480  
QY 106 WNAKN-----SDITVGYDQDLVTRV 125  
DB 481 WGVSNFTGRTISVTVNGSGTAVTTI 505

RESULT 14  
US-60-565-632-9194  
Sequence 9194, Application US/60565632  
GENERAL INFORMATION:  
APPLICANT: Monsanto Technology, LLC  
APPLICANT: Baum, James A  
APPLICANT: Kovalic, David K.  
APPLICANT: Larosa, Thomas J.  
APPLICANT: Lu, Maolong  
APPLICANT: Munyikwa, Tichifa R. I.  
APPLICANT: Roberts, James K.  
APPLICANT: Wu, Wei  
APPLICANT: Zhang, Bei  
TITLE OF INVENTION: Methods for Genetic Control of Insect Infestations in Plants and  
TITLE OF INVENTION: Compositions thereof  
FILE REFERENCE: 38-21 (53403)B  
CURRENT FILING DATE: 2004-04-27  
CURRENT APPLICATION NUMBER: US/60/565,632  
NUMBER OF SEQ ID NOS: 15449  
SOFTWARE: Patent in version 3.2  
SEQ ID NO 9194  
LENGTH: 758  
TYPE: PRT  
ORGANISM: Diabrotica virgifera  
US-60-565-632-9194

Query Match 10.6%; Score 82; DB 7; Length 758;  
Best Local Similarity 26.8%; Pred. No. 47;  
Matches 33; Conservative 21; Mismatches 41; Indels 28; Gaps 7;  
QY 29 GGNHGGGNSGPDSTLSIYQYGSANAALQSDARKSETTITQSGYNGADVGGADNS 88  
DB 77 GHENNGHYGSSSTKTITQTSSS-----SQGNLEQITL-NHGNNG-IGYSGSK 128  
QY 89 TIELTQNGFRNATIDQNAKNSDITVGYDQDLVTRVTHE-MAHASVMVRQVGFNNAT 147  
DB 129 VIE-----TTVTKHLGGNNEKHLGKLDQV--LTGHENNGH-----GYGSSST 169  
QY 148 ANQ 150  
DB 170 KTQ 172

Search completed: August 2, 2004, 15:29:51  
Job time : 18.8 secs

APPLICANT: Diversa Corporation  
APPLICANT: Steer, Brian  
APPLICANT: Callen, Walter  
APPLICANT: Healey, Shaun  
APPLICANT: Pulliam, Derrick  
TITLE OF INVENTION: GLUCANASES, NUCLEIC ACIDS ENCODING THEM AND METHODS FOR MAKING AN  
TITLE OF INVENTION: THEM  
FILE REFERENCE: 564462009540  
CURRENT APPLICATION NUMBER: PCT/US04/21492  
CURRENT FILING DATE: 2004-07-20  
PRIOR APPLICATION NUMBER: 60/484,725  
PRIOR FILING DATE: 2003-07-02  
NUMBER OF SEQ ID NOS: 518  
SOFTWARE: Patent in version 3.1  
SEQ ID NO 88  
LENGTH: 535  
TYPE: PRT  
ORGANISM: Unknown  
FEATURE:  
OTHER INFORMATION: Obtained from an environmental sample.  
NAME/KEY: SIGNAL  
LOCATION: (1)...(27)  
PCT-US04-21492-88

Query Match 10.6%; Score 82.5; DB 1; Length 535;  
Best Local Similarity 26.8%; Pred. No. 29;  
Matches 38; Conservative 18; Mismatches 58; Indels 31; Gaps 7;  
QY 12 IVVSG-----SALAGVVP--QWGGGN-----HNGG-----GNSGGPDSTLSIYQYGA 53  
DB 361 IVVNGTQRAENGSVNTGWANNQCGSGNSELHCGNVIISFGVSGSSSSSSSSSSSSSS 420  
QY 54 NAALQSDARKSETTITQ-----SGVNGADVGGADNSTIELT-QNGFRRNATIQQ 105  
DB 421 SSSSTSS 480  
QY 106 WNAKN-----SDITVGYDQDLVTRV 125  
DB 481 WGVSNFTGRTISVTVNGSGTAVTTI 505

RESULT 14  
US-60-565-632-9194  
Sequence 9194, Application US/60565632  
GENERAL INFORMATION:  
APPLICANT: Monsanto Technology, LLC  
APPLICANT: Baum, James A  
APPLICANT: Kovalic, David K.  
APPLICANT: Larosa, Thomas J.  
APPLICANT: Lu, Maolong  
APPLICANT: Munyikwa, Tichifa R. I.  
APPLICANT: Roberts, James K.  
APPLICANT: Wu, Wei  
APPLICANT: Zhang, Bei  
TITLE OF INVENTION: Methods for Genetic Control of Insect Infestations in Plants and  
TITLE OF INVENTION: Compositions thereof  
FILE REFERENCE: 38-21 (53403)B  
CURRENT FILING DATE: 2004-04-27  
CURRENT APPLICATION NUMBER: US/60/565,632  
NUMBER OF SEQ ID NOS: 15449  
SOFTWARE: Patent in version 3.2  
SEQ ID NO 9194  
LENGTH: 758  
TYPE: PRT  
ORGANISM: Diabrotica virgifera  
US-60-565-632-9194

Query Match 10.6%; Score 82; DB 7; Length 758;  
Best Local Similarity 26.8%; Pred. No. 47;  
Matches 33; Conservative 21; Mismatches 41; Indels 28; Gaps 7;  
QY 29 GGNHGGGNSGPDSTLSIYQYGSANAALQSDARKSETTITQSGYNGADVGGADNS 88  
DB 77 GHENNGHYGSSSTKTITQTSSS-----SQGNLEQITL-NHGNNG-IGYSGSK 128  
QY 89 TIELTQNGFRNATIDQNAKNSDITVGYDQDLVTRVTHE-MAHASVMVRQVGFNNAT 147  
DB 129 VIE-----TTVTKHLGGNNEKHLGKLDQV--LTGHENNGH-----GYGSSST 169  
QY 148 ANQ 150  
DB 170 KTQ 172

Search completed: August 2, 2004, 15:29:51  
Job time : 18.8 secs



GenCore version 5.1.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: August 2, 2004, 14:39:53 ; Search time 9.4 Seconds  
(without alignments)  
1545.204 Million cell updates/sec

Title: US-09-543-407-14  
Perfect score: 775  
Sequence: 1 MKLLKVAFAAIVVSGSALA.....HASVMVRQVGFNNATANQY 151

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 283366 seqs, 96191526 residues

Total number of hits satisfying chosen parameters: 283366

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : PIR 78.\*  
1: Piri.\*  
2: Piri2.\*  
3: Piri3.\*  
4: Piri4.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Match	Length	DB ID	Description
1	693	89.4	151	2	JC6039
2	693	89.4	151	2	fimbrin protein ag
3	528	68.1	151	2	major curlin chain
4	506.5	65.4	152	2	curlin protein csg
5	506.5	65.4	152	2	curlin major subun
6	104.5	13.5	151	2	hypothetical prote
7	104.5	13.5	151	2	curlin nucleator p
8	104.5	13.5	151	2	minor curlin subun
9	98.5	12.7	151	2	curlin minor chain
10	96.5	12.5	151	2	fimbrin protein ag
11	96	12.4	770	2	probable PPE prote
12	95	12.3	1748	2	related to C2H2 zi
13	93	12.0	2174	2	cnjB protein - Tet
14	91	11.7	301	2	hypothetical glyci
15	91	11.7	1028	2	hypothetical protei
16	91	11.7	1213	2	DNA-binding protei
17	90.5	11.7	145	2	ovo protein - fru
18	90.5	11.7	145	2	conserved hypotet
19	90.5	11.7	151	2	hypothetical prote
20	90	11.6	479	2	nucleation compone
21	89.5	11.5	256	2	metalloproteinase
22	89.5	11.5	262	2	glycine-rich prote
23	89	11.5	573	2	Tail fiber protein
24	88.5	11.4	479	2	F3f19.21 protein -
25	87	11.2	590	2	metalloproteinase
26	87	11.2	646	1	probable PPE prote
27	87	11.2	678	2	leishmanolysin (EC
28	86.5	11.2	478	2	probable PPE prote
29	86.5	11.2	527	2	metalloproteinase
					hypothetical prote

30 86.5 11.2 645 2 F70825 probable PPE prote  
31 86 11.1 447 2 G84687 probable disease r  
32 86 11.1 599 2 B42049 leishmanolysin (EC  
33 86 11.1 599 2 A44951 leishmanolysin (EC  
34 85.5 11.0 575 2 S35327 protein kinase sgg  
35 85 11.0 347 2 B39112 merocaine 45K surf  
36 85 11.0 967 2 S66852 hypochetrical prote  
37 84.5 10.9 440 2 AD1539 probable sugar ABC  
38 84 10.8 407 2 T21956 leishmanolysin (EC  
39 84 10.8 590 1 A45621 leishmanolysin (EC  
40 84 10.8 602 1 PL0221 hypochetrical prote  
41 83.5 10.8 285 2 H84219 hypochetrical prote  
42 83.5 10.8 340 2 A83401 hypochetrical prote  
43 83.5 10.8 440 2 T15352 hypochetrical prote  
44 83.5 10.8 1067 2 S35423 protein kinase sgg  
45 83.5 10.8 3300 2 D70575 probable PPE prote

## ALIGNMENTS

## RESULT 1

JC6039  
fimbrin protein agfa precursor - Salmonella enteritidis  
C:Species: Salmonella enteritidis  
C:Date: 31-Dec-1996 #sequence: revision 31-Dec-1996 #text\_change 08-Oct-1999  
C:Accession: JC6039; PC6015; A44898  
R:Collinson, S.K.; Clouthier, S.C.; Doran, J.L.; Baner, P.A.; Kay, W.W.  
J. Bacteriol. 178, 662-667, 1996  
A:Title: Salmonella enteritidis agfBAC operon encoding thin, aggregative fimbriae.  
A:Reference number: JC6039; MUID:96146512; PMID:8550497  
A:Accession: JC6039  
A:Molecule type: DNA  
A:Residues: 1-151 <COL>  
A:Cross-references: GB:U43280; NID:gl184712; PIDN:AAC43599.1; PID:gl184714  
A:Accession: PC6015  
A:Molecule type: Protein  
A:Residues: 21-52 <CO2>  
A:Experimental source: strain 27655-3b  
A:Note: The authors translated the codon ACG for residue 44 as Ile  
R:Collinson, S.K.; Emody, L.; Muller, K.H.; Trust, T.J.; Kay, W.W.  
J. Bacteriol. 173, 4773-4781, 1991  
A:Title: Purification and characterization of thin, aggregative fimbriae from Salmonella  
A:Reference number: A44898; MUID:91310586; PMID:1677357  
A:Contents: 27655  
A:Accession: A44898  
A>Status: preliminary  
A:Molecule type: protein  
A:Residues: 21-33 <CO3>  
A:Note: sequence extracted from NCBI backbone (NCBIP:45936)  
C:Genetics:  
C:Gene: agfa  
C:Function:  
A:Description: major component of thin aggregative fimbriae  
A:Note: fimbriae bind to fibronectin, plasminogen, tissue plasminogen activator  
F:1-20/Domain: signal sequence #status predicted <SIG>  
F:21-151/Product: fimbrin protein agfa #status experimental <MAT>

Query Match 89.4%; Score 693; DB 2; Length 151;  
Best Local Similarity 90.7%; Pred. No. 58-52; Indels 0; Gaps 0;  
Matches 137; Conservative 3; Mismatches 11;

Qy 1 MKLLKVAFAAIVVSGSALAAGVVPQWGGGNGNGSSGPDSTLSIYQYGSANAALALQ 60  
Db 1 MKLLKVAFAAIVVSGSALAAGVVPQWGGGNGNGSSGPDSTLSIYQYGSANAALALQ 60  
Qy 61 SDARKSETTITQSGYNGADVQGGADNSETIETONGFRNNATIDOWNAKNSDITVGYQGG 120  
Db 61 SDARKSETTITQSGYNGADVQGGADNSETIETONGFRNNATIDOWNAKNSDITVGYQGG 120  
Qy 121 LVTRVVTHEMAHASVMVRQVGFNNATANQY 151





```
Db 61 QADARNSDLTITQHGGGNGADVGQSDSSIDLTQRFNGSATLDQWNGKDSHMTVKQFG 120
QY 120 QLVTRVVTHEMAHASVWVRQVGFNNATANQY 151
Db 121 GGNGAAVDQTASNTVNVTVQVGFNNATANQY 152
RESULT 5
H85665
hypothetical protein csgA [imported] - Escherichia coli (strain O157:H7, substrain EDL93)
C:Species: Escherichia coli
C>Date: 16-Feb-2001 #sequence_revision 16-Feb-2001 #text_change 14-Sep-2001
C:Accession: H85665
R:Perna, N.T.; Plunkett III, G.; Burland, V.; Mau, B.; Glasner, J.D.; Rose, D.J.; Mayhew,
iller, L.; Grobeck, E.J.; Davis, N.W.; Lim, A.; Dimalanta, E.; Potamousis, K.; Apodaca,
Nature 409, 529-533, 2001
A:Title: Genome sequence of enterohemorrhagic Escherichia coli O157:H7.
A:Reference number: A85480; MUID:21074935; PMID:11206551
A:Accession: H85665
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-152 <STO>
A:Cross-references: GB:AE005174; NID:g12514574; PIDN:AA055788.1; GSPDB:GN00145; UWGP:216
A:Experimental source: strain O157:H7, substrain EDL933
C:Genetics:
A:Gene: csgA
Query Match 65.4%; Score 506.5; DB 2; Length 152;
Best Local Similarity 67.1%; Pred. No. 3.6e-36;
Matches 102; Conservative 21; Mismatches 28; Indels 1; Gaps 1;
QY 1 MKLLKVAAPAAIVVSGSALAGVVPQW-GGGGNGGNGSGSPDSTLSIYQYGSANAALAL 59
Db 1 MKLLKVAALAAIVFSGSALAGVVPQYGGGNGGNGSGSPSELNIYQYGGGNSALAL 60
QY 60 QSDARKSETTITQSGYNGADVGQADNSTIELTQNGFRNNATIDQWNAKNSDITVGYQVD 119
Db 61 QADARNSDLTITQHGGGNGADVGQSDSSIDLTQRFNGSATLDQWNGKDSHMTVKQFG 120
QY 120 QLVTRVVTHEMAHASVWVRQVGFNNATANQY 151
Db 121 GGNGAAVDQTASNTVNVTVQVGFNNATANQY 152
RESULT 6
S70787
curlin nucleator protein csgB precursor - Escherichia coli (strain K-12)
N:Alternate names: csgB protein; curlin nucleation component; minor curlin protein
C:Species: Escherichia coli
C>Date: 12-Feb-1998 #sequence_revision 20-Feb-1998 #text_change 01-Mar-2002
C:Accession: S70787; F64846
R:Hammar, M.; Arnqvist, A.; Bian, Z.; Olsen, A.; Normark, S.
Mol. Microbiol. 18, 661-670, 1995
A:Title: Expression of two csg operons is required for production of fibronectin- and Cd
A:Reference number: S70783; MUID:96414468; PMID:8817489
A:Accession: S70787
A:Status: nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-151 <HAM>
A:Cross-references: ENBL:X90754; NID:g1147558; PIDN:CAA62281.1; PID:g1147563
A:Experimental source: strain K12, substrain W3110
A:Note: the nucleotide sequence was submitted to the EMBL Data Library, August 1995
R:Blattner, F.R.; Plunkett III, G.; Bloch, C.A.; Perna, N.T.; Burland, V.; Riley, M.; Co
A.; Rose, D.J.; Mau, B.; Shao, Y.
Science 277, 1453-1462, 1997
A:Title: The complete genome sequence of Escherichia coli K-12.
A:Reference number: A64720; MUID:97426617; PMID:9278503
A:Accession: F64846
A:Status: nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-151 <BLAT>
A:Cross-references: GB:AE000205; GB:U00096; NID:g1787265; PIDN:AA074125.1; PID:g1787278;
```

```
A:Experimental source: strain K-12, substrain MG1655
C:Genetics:
A:Gene: csgB
A:Map position: 23.15
C:Function:
A:Description: minor component of wild-type curli; interaction between CsgA and CsgB triic
A:Note: Curli are thin, coiled fibers expressed on the surface of Escherichia coli that i
and H-kininogen; in the absence of CsgA, CsgB can self-assemble into polymers
F:1-21/Domain: signal sequence #status predicted <SIG>
F:122-151/Product: minor curlin chain #status predicted <MAT>
Query Match 13.5%; Score 104.5; DB 2; Length 151;
Best Local Similarity 29.2%; Pred. No. 0.054;
Matches 35; Conservative 19; Mismatches 59; Indels 7; Gaps 4;
QY 9 FAAIVVSGSAL--AGVVPQWGGGNGGNGSGSPDSTLSIY-QYGSANAALALQSDARK 65
Db 33 FAVNELSKSFPNQAAII---GQAGTNNSAQLRQGGSKLLAVVAQEGSSNRA-KIDQTGDY 88
QY 66 SETTITQSGYNGADVGQADNSTIELTQNGFRNNATIDQWNAKNSDITVGYQDQLVTRV 125
Db 89 NLAYIDQAGSANDASISQAGYNTAMIIQKSGNKANITQYGTQKTAIVVQRQSQWAIKV 148
RESULT 7
C90806
minor curlin subunit precursor CsgB [imported] - Escherichia coli (strain O157:H7, substi
C:Species: Escherichia coli
C>Date: 18-Jul-2001 #sequence_revision 18-Jul-2001 #text_change 18-Jul-2001
C:Accession: C90806
R:Hayashi, T.; Makino, K.; Ohnishi, M.; Kurokawa, K.; Ishii, K.; Yokoyama, K.; Han, C.G.;
gasawara, N.; Yasunaga, T.; Kuhara, S.; Shiba, T.; Hattori, M.; Shinagawa, H.
DNA Res. 8, 11-22, 2001
A:Title: Complete genome sequence of enterohemorrhagic Escherichia coli O157:H7 and genom
A:Reference number: A99629; MUID:21156231; PMID:11258796
A:Accession: C90806
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-151 <HAY>
A:Cross-references: GB:BA000007; PIDN:BA034842.1; PID:g13360879; GSPDB:GN00154
A:Experimental source: strain O157:H7, substrain RIMD 0509952
C:Genetics:
A:Gene: SC51419
Query Match 13.5%; Score 104.5; DB 2; Length 151;
Best Local Similarity 29.2%; Pred. No. 0.054;
Matches 35; Conservative 19; Mismatches 59; Indels 7; Gaps 4;
QY 9 FAAIVVSGSAL--AGVVPQWGGGNGGNGSGSPDSTLSIY-QYGSANAALALQSDARK 65
Db 33 FAVNELSKSFPNQAAII---GQAGTNNSAQLRQGGSKLLAVVAQEGSSNRA-KIDQTGDY 88
QY 66 SETTITQSGYNGADVGQADNSTIELTQNGFRNNATIDQWNAKNSDITVGYQDQLVTRV 125
Db 89 NLAYIDQAGSANDASISQAGYNTAMIIQKSGNKANITQYGTQKTAIVVQRQSQWAIKV 148
RESULT 8
G85665
curlin minor chain precursor, CsgA homolog [imported] - Escherichia coli (strain O157:H7,
C:Species: Escherichia coli
C>Date: 16-Feb-2001 #sequence_revision 16-Feb-2001 #text_change 14-Sep-2001
C:Accession: G85665
R:Perna, N.T.; Plunkett III, G.; Burland, V.; Mau, B.; Glasner, J.D.; Rose, D.J.; Mayhew,
iller, L.; Grobeck, E.J.; Davis, N.W.; Lim, A.; Dimalanta, E.; Potamousis, K.; Apodaca,
Nature 409, 529-533, 2001
A:Title: Genome sequence of enterohemorrhagic Escherichia coli O157:H7.
A:Reference number: A85480; MUID:21074935; PMID:11206551
A:Accession: G85665
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-151 <STO>
A:Cross-references: GB:AE005174; NID:g12514573; PIDN:AA055787.1; GSPDB:GN00145; UWGP:216;
```

A;Experimental source: strain O157:H7, substrain EDL933  
 C;Genetics:  
 A;Gene: cs9B

Query Match 13.5%; Score 104.5; DB 2; Length 151;  
 Best Local Similarity 29.2%; Pred. No. 0.054;  
 Matches 35; Conservative 19; Mismatches 59; Indels 7; Gaps 4;

QY 9 FAATVVGSSAL--AGVVPQGGGNGHNGSGSPDSTLSIY-QYGSANAALALQSDARK 65

DB 33 FAVNELSKSFNQRAII---GQAGTNSAQ:RQGGSKLLAVAQEGSSNRA-KIDQTGY 88

QY 66 SETTITQSGYNGADVGQGGADNSTIETQNGFRNATIDQNAKNSDITVGYDQLVTR 125

DB 89 NLAVIDQAGSANDASISQAGYGNNTAMIIQKSGNKANITQYGTQTAIVVQROQMAIRV 148

# RESULT 9

JC6040

fimbria protein agfB precursor - Salmonella enteritidis

C;Species: Salmonella enteritidis

C;Date: 31-Dec-1996 #sequence\_revision 31-Dec-1996 #text\_change 08-Oct-1999

C;Accession: JC6040

R;Collinson, S.K.; Clouthier, S.C.; Doran, J.L.; Baner, P.A.; Kay, W.W.

J. Bacteriol. 178, 662-667, 1996

A;Title: Salmonella enteritidis agfBAC operon encoding thin, aggregative fimbriae.

A;Reference number: JC6039; MUID:96146512; PMID:8550497

A;Accession: JC6040

A;Molecule type: DNA

A;Residues: 1-151 <COL>

A;Cross-references: GB:U43280; NID:g1184712; PIDN:AAC43598.1; PID:g1184713

A;Experimental source: strain 276755-3b

C;Genetics:

A;Gene: agfB

C;Function:

A;Description: minor component of thin aggregative fimbriae

A;Note: fimbriae bind to fibronectin, plasminogen, tissue plasminogen activator

C;Keywords: fimbria

F;1-21/Domain: signal sequence #status predicted <SIG>

F;22-151/Product: fimbria protein agfB #status predicted <MAT>

Query Match 12.7%; Score 98.5; DB 2; Length 151;  
 Best Local Similarity 26.3%; Pred. No. 0.17;  
 Matches 34; Conservative 17; Mismatches 42; Indels 37; Gaps 4;

QY 10 AATVVGSSALAGVVPQGGGNGHNGSGSPDSTLSIYQYGSANAALALQSDARKSETT 69

DB 58 ARVRQEGSKLLSVISQ--EGGNRAKVDQAGNYNFAYIEQTGNAN-----DAS 103

QY 70 ITQSGYNGADVGQGGADNSTIETQNGFRNATIDQNAKNSDITVGYDQLVTRVWTHE 129

DB 104 ISQSAVGNLSA-----AIIQKSGNKANITQYGTQK-----TAVVQX 140

QY 130 MAHASVWVRQ 139

DB 141 QSHMAIRVTQ 150

# RESULT 10

D70604

probable PPE protein - Mycobacterium tuberculosis (strain H37Rv)

C;Species: Mycobacterium tuberculosis

C;Date: 17-Jul-1998 #sequence\_revision 17-Jul-1998 #text\_change 20-Jun-2000

C;Accession: D70604

R;Cole, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon, S.

Connor, R.; Davies, R.; Devlin, K.; Feltwell, T.; Gentles, S.; Hamlin, N.; Holroyd, S.

Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S.

Nature 393, 537-544, 1998

A;Authors: Sgares, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.

A;Title: Deciphering the biology of Mycobacterium tuberculosis from the complete genome

A;Reference number: A70500; MUID:98295987; PMID:9634230

A;Accession: D70604

A;Status: preliminary; nucleic acid sequence not shown; translation not shown

A;Molecule type: DNA

A;Residues: 1-552 <COL>

A;Cross-references: GB:Z92774; GB:AL123456; NID:g3261729; PIDN:CAB07133.1; PID:g1877289

A;Experimental source: strain H37Rv

C;Genetics:

A;Gene: PPE

C;Superfamily: Phaseolus glycine-rich cell wall protein 1.8

Query Match 12.5%; Score 96.5; DB 2; Length 552;  
 Best Local Similarity 28.5%; Pred. No. 1.1;  
 Matches 41; Conservative 10; Mismatches 62; Indels 31; Gaps 6;

QY 15 SGSALAGV-----VPQGGGNGHNGSGSPDSTLSIYQYGSANAALALQSDA 63

DB 353 SGSGNIGVNTGANTLPV---GDLNLGVGSGNANITGFGVGLNTGFGNASILNTGLG 409

QY 64 RKSETTITQSGYNGADVGQGGADNSTIETQNGFRNATIDQNAKNSDITVGYDQLVT 123

DB 410 NAGE---LNTGFGNAGVNTGFDNSGNVNTGNGSGNINTGSMNAGNVNTGFG----- 459

QY 124 RVVTHEMAHASVWVRQVQFGNNAT 147

DB 450 -IITDSGLTNS-----GFGNTGT 476

# RESULT 11

T51024

N;Alternate names: protein B7F21.50

C;Species: Neurospora crassa

C;Date: 21-Jul-2000 #sequence\_revision 21-Jul-2000 #text\_change 21-Jul-2000

C;Accession: T51024

R;Schulte, U.; Aign, V.; Hoheisel, J.; Brandt, P.; Fartmann, B.; Holland, R.; Nyakatura,

submitted to the Protein Sequence Database, July 2000

A;Reference number: Z25286

A;Accession: T51024

A;Status: preliminary

A;Molecule type: DNA

A;Residues: 1-770 <SCH>

A;Cross-references: EMBL:AL389901; GSPDB:GN00116; NCSP:B7F21.50

A;Experimental source: BAC clone B7F21; strain OR74A

C;Genetics:

A;Gene: NCSP:B7F21.50

A;Map position: 6

A;Inserts: 117/1

Query Match 12.4%; Score 96; DB 2; Length 770;  
 Best Local Similarity 22.8%; Pred. No. 1.8;  
 Matches 31; Conservative 19; Mismatches 54; Indels 32; Gaps 3;

QY 28 GGGNHNGSGSGSPDSTLSIYQYGSANAALALQSDARKSETTITQSGYNGADVGQADN 87

DB 389 GAGNNNGSELASDPNRYFD---GASPLPSSSTTTTAYAGHNGHNGHRESVDS 444

QY 88 STIEL-----TQNGFRNATIDQNAKNSDITVGYDQLV----- 122

DB 445 QSTGLGVHYSSTTRNQHQHWRNQSNASLSADGSEITHGVASPLVGGSSHARGSGGTS 504

QY 123 ---TRVVTHEMAHASV 135

DB 505 YRYTHTHSHSHSGL 520

# RESULT 12

S42136

cnjB protein - Tetrahymena thermophila

C;Species: Tetrahymena thermophila

C;Date: 19-Mar-1997 #sequence\_revision 18-Jul-1997 #text\_change 07-Dec-1999

C;Accession: S42136; S42135; S03650

R;Taylor, F.M.; Martindale, D.W.

submitted to the EMBL Data Library, October 1992

A;Reference number: S42136

A;Accession: S42136

A:Molecule type: DNA  
 A:Residues: 1-1748 <TAY>  
 A:Cross-references: EMBL:L03710; NID:g161751; PID:g161752  
 R:Taylor, F.M.; Martindale, D.W.  
 Nucleic Acids Res. 21, 4610-4614, 1993  
 A:Title: Retroviral-type zinc fingers and glycine-rich repeats in a protein encoded by c  
 A:Reference number: S42135; MUID:94051569; PMID:8233798  
 A:Accession: S42135  
 A:Molecule type: DNA  
 A:Residues: 1164-1174; 1179-1198; 1233-1252; 1285-1293; 1297-1309; 1316-1326; 1331-1341; 1343-1  
 A:Cross-references: EMBL:L03710  
 R:Martindale, D.W.; Taylor, F.M.  
 Nucleic Acids Res. 16, 2189-2201, 1988  
 A:Title: Multiple introns in a conjugation-specific gene from *Tetrahymena thermophila*.  
 A:Reference number: S03650; MUID:88189811; PMID:3357771  
 A:Accession: S03650  
 A:Molecule type: DNA  
 A:Residues: 236-250, '1', 252-255, 'N', 257-773 <MAR>  
 A:Cross-references: EMBL:X06482  
 C:Genetics:  
 A:Gene: *cnjB*  
 A:Genetic code: SGC5  
 A:Introns: 85/3; 136/1; 157/3; 201/2; 290/2; 327/3; 499/1; 573/2; 607/3; 708/3; 777/3; 8  
 C:Keywords: zinc finger  
 F:1164-1450/Region: Glycine-rich  
 F:1451-1484/Region: zinc finger CCHC motif  
 F:1478-1491/Region: zinc finger CCHC motif  
 F:1501-1514/Region: zinc finger CCHC motif  
 F:1530-1543/Region: zinc finger CCHC motif  
 F:1555-1568/Region: zinc finger CCHC motif  
 F:1579-1592/Region: zinc finger CCHC motif  
 F:1602-1615/Region: zinc finger CCHC motif  
 F:1626-1748/Region: glycine-rich

Query Match 12.3%; Score 95; DB 2; Length 1748;  
 Best Local Similarity 32.0%; Pred. No. 5.6;  
 Matches 31; Conservative 11; Mismatches 29; Indels 26; Gaps 4;  
 A:Map position: 2

QY 25 OWGGGNGHGG--GNSGDPSTLSIYQGSANAALALQSDARKSETTIT---QSGYGN 77  
 DB 1640 QFGGGGNSGGQSGWTSGSDWN-----CQSNVQESITTSGGWSSGSGN 1685  
 QY 78 GADYGGGADNSTIELTQNGFRNATIDQWNAKNSDIT 114  
 DB 1686 QTGGGWSGNDN-----QQQCNENTGGGWSGSSNSQT 1717

RESULT 13  
 B95965  
 hypothetical glycine-rich protein [imported] - *Sinorhizobium meliloti* (strain 1021) mega  
 C:Species: *Sinorhizobium meliloti*  
 C:Date: 24-Aug-2001 #sequence\_revision 24-Aug-2001 #text\_change 30-Sep-2001  
 C:Accession: B95965  
 R:Finan, T.M.; Weidner, S.; Wong, K.; Buhrmester, J.; Chain, P.; Vorholter, F.J.; Hernar  
 Proc. Natl. Acad. Sci. U.S.A. 96, 9889-9894, 2001  
 A:Title: The complete sequence of the 1,683-Kb pSymB megaplasmid from the N2-fixing endo  
 A:Reference number: A95842; MUID:21396508; PMID:11481431  
 A:Accession: B95965  
 A:Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-2174 <KUR>  
 A:Cross-references: GB:ALU591985; PID:CA49389.1; PID:g15140875; GSPDB:GN00167  
 A:Experimental source: strain 1021, megaplasmid pSymB  
 R:Galibert, F.; Finan, T.M.; Long, S.R.; Puhler, A.; Abola, P.; Ampe, F.; Barloy-Hubler,  
 Pella, D.; Chain, P.; Cowie, A.; Davis, R.W.; Dreano, S.; Federspiel, N.A.; Fisher, R.F.;  
 L.; Hyman, R.W.; Jones, T.  
 Science 293, 668-672, 2001  
 A:Authors: Kahn, D.; Kahn, M.L.; Kalman, S.; Keating, D.H.; Kiss, E.; Komp, C.; Lelaure,  
 hebaull, P.; Vandenbol, M.; Vorholter, F.J.; Weidner, S.; Wells, D.H.; Wong, K.; Yeh, K.  
 A:Title: The composite genome of the legume symbiont *Sinorhizobium meliloti*.  
 A:Reference number: A96039; MUID:21368234; PMID:11474104  
 A:Contents: annotation  
 C:Genetics:

A:Gene: SMD21548  
 A:Genome: plasmid

Query Match 12.0%; Score 93; DB 2; Length 2174;  
 Best Local Similarity 24.5%; Pred. No. 11;  
 Matches 36; Conservative 15; Mismatches 32; Indels 64; Gaps 7;  
 QY 11 AIVVSGSALAGVVPQ--WGCGGNGHNGGNSGDPSTLSIYQGSANAALALQSDARKSET 68  
 DB 693 AATAGAGAVGILAQSTGGGNG--CGNATGGDAGFGSGFIQGGGGG----- 737  
 QY 69 TITQSGYNGADVG--QGADNSTIELTQNGFRNATIDQWNAKNSDITVQYDQLVTRVVT 127  
 DB 738 ---GGYANTANVFGK-----LTLTTQG----- 757  
 QY 128 HEMAHASVMVQV---GFGNATANQY 151  
 DB 758 ---SHAAGIVAQSVGGGGTGTGTASSY 781

## RESULT 14

B84533  
 hypothetical protein At2g15770 [imported] - *Arabidopsis thaliana*  
 C:Species: *Arabidopsis thaliana* (mouse-ear cress)  
 C:Date: 02-Feb-2001 #sequence\_revision 02-Feb-2001 #text\_change 02-Feb-2001  
 C:Accession: B84533  
 R:Lin, X.; Kaul, S.; Rounsley, S.D.; Shea, T.P.; Benito, M.I.; Town, C.D.; Fujii, C.Y.; A  
 M.; Koo, H.; Moffat, K.S.; Cronin, L.A.; Shen, M.; VanAken, S.E.; Umayam, L.; Tallon, L.  
 euss, D.; Nierman, W.C.; White, O.; Eisen, J.A.; Salzberg, S.L.; Fraser, C.M.; Venter, J.  
 Nature 402, 761-768, 1999  
 A:Title: Sequence and analysis of chromosome 2 of the plant *Arabidopsis thaliana*.  
 A:Reference number: A84420; MUID:20083487; PMID:10617197  
 A:Accession: B84533  
 A:Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-301 <STO>  
 A:Cross-references: GB:AE002093; NID:g5306254; PID:AD41987.1; GSPDB:GN00139  
 C:Genetics:  
 A:Gene: At2g15770  
 A:Map position: 2

Query Match 11.7%; Score 91; DB 2; Length 301;  
 Best Local Similarity 26.4%; Pred. No. 1.7;  
 Matches 33; Conservative 19; Mismatches 51; Indels 22; Gaps 5;

QY 15 SCSALA-----GVVPQWGGCGGNH--NGGNSGDPSTLSIYQGSANAALALQSDARKSET 68  
 DB 69 SGGGLGNSNNGSGWGTGSGSGSTNPDGRRSRWNW-----SLKSGWSWSWG 121

QY 69 TITQSGYNGADVGOGADNST---IELTQNGFRNATIDQWNAKNSDITVQYD 119  
 DB 122 SNDNSNSGSDSGSLDRETPKIIVGGSDGKGLDYKDWASKNAPFYVNDVLVFKYD 181

QY 120 QLVTR 124

DB 182 KSAKR 186

## RESULT 15

A56038  
 DNA-binding protein ovo - fruit fly (*Drosophila melanogaster*)  
 C:Species: *Drosophila melanogaster*  
 C:Date: 01-Dec-1995 #sequence\_revision 01-Dec-1995 #text\_change 21-Jul-2000  
 C:Accession: A56038  
 R:Garfield, M.D.; Wang, J.; Liang, Y.; Mahowald, A.P.  
 Mol. Cell. Biol. 14, 6809-6818, 1994  
 A:Title: Multiple products from the shavenbaby-ovo gene region of *Drosophila melanogaster*  
 A:Reference number: A56038; MUID:95021209; PMID:7935398  
 A:Accession: A56038  
 A:Status: preliminary  
 A:Molecule type: mRNA  
 A:Residues: 1-1028 <GAR>  
 A:Cross-references: GB:U11383; NID:g520526; PID:NAAB60216.1; PID:g520527

C:Genetics:  
A:Gene: ovo  
A:Cross-references: FlyBase:FBgn0003028  
Query Match 11.7%; Score 91; DB 2; Length 1028;  
Best Local Similarity 31.2%; Pred. No. 6.7;  
Matches 34; Conservative 8; Mismatches 41; Indels 26; Gaps 4;  
Qy 3 LLKVAAPAAIVVSGSALAGVVPQWGGGNNHGGGNSGDPSTLSIYCYGSANAALALQSD 62  
Db 59 LQNAARAAIYMSAGS-----GGCTGNGGGGSGPGGFSANSGGGGGG----- 104  
Qy 63 ARKSETTITQSGYNGADVQGGADNSTIELTQNGFRNNATIDQNAKNS 111  
Db 105 -----GGNGYINCGVG-GPNNS---LDGNLLNFASVSNYESNS 141

Search completed: August 2, 2004, 14:56:21  
Job time : 9.4 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: August 2, 2004, 14:36:12 ; Search time 5.3 Seconds  
(without alignments)

1483.508 Million cell updates/sec

Title: US-09-543-407-14

Perfect score: 775  
Sequence: 1 MKLLKVAFAAIVVSGSALA.....HASVMVRQVGFNNATANQY 151

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 141681 seqs, 52070155 residues

Total number of hits satisfying chosen parameters: 141681

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : SwissProt\_42.\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	693	89.4	151	1	CSGA_SALTY
2	528	68.1	151	1	CSGA_ECOLI
3	506.5	65.4	152	1	CSGA_ECO57
4	104.5	13.5	151	1	CSGB_ECOLI
5	98.5	12.7	151	1	CSGB_SALTY
6	92	11.9	1327	1	TNKK_HUMAN
7	91	11.7	1028	1	OVO_DROME
8	90.5	11.7	151	1	CSGB_SALTY
9	90	11.6	1093	1	VG38_BPT2
10	89.5	11.5	262	1	PER_DROWI
11	88.5	11.4	479	1	PRTC_ERWCH
12	87	11.2	646	1	GP63_LEIME
13	87	11.2	678	1	YF48_MYCTU
14	86	11.1	599	1	GP63_LEICH
15	85.5	11.0	1048	1	AGOI_ARATH
16	85	11.0	347	1	MSA2_PLAF2
17	84	10.8	590	1	GP63_LEIDO
18	84	10.8	602	1	GP63_LEIMA
19	83.5	10.8	440	1	PDAG_CAEEL
20	83.5	10.7	362	1	SGG_DROME
21	83	10.7	1656	1	P35_MYCPE
22	83	10.7	1656	1	OMPB_RICJA
23	82	10.6	1080	1	HDC_DROME
24	81.5	10.5	392	1	HMEI_HUMAN
25	81	10.5	165	1	GRPI_ORYSA
26	81	10.5	1612	1	RRPO_PMVUJ
27	81	10.5	1612	1	RRPO_PMVVS
28	80.5	10.4	481	1	PRTB_ERWCH
29	80.5	10.4	641	1	IMD_ARTGO
30	80	10.3	385	1	PER_DROME
31	80	10.3	491	1	YK98_MYCTU
32	80	10.3	548	1	CEAK_ECOLI
33	79.5	10.3	204	1	HEVE_HEVER

ALIGNMENTS

RESULT 1

ID	CSGA_SALTY	STANDARD;	PRT;	151 AA.
AC	P55225;			
DT	01-OCT-1996 (Rel. 34, Created)			
DT	01-OCT-1996 (Rel. 34, Last sequence update)			
DT	10-OCT-2003 (Rel. 42, Last annotation update)			
DE	Major curlin subunit precursor (Pimbrin SBF17).			
GN	CSGA OR AGFA OR STM1144 OR STV1181 OR T1776.			
OS	Salmonella typhimurium,			
OS	Salmonella typhi, and			
OC	Salmonella enteritidis.			
OC	Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;			
OC	Enterobacteriaceae; Salmonella.			
OX	NCBI_TaxID=602, 601, 592;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RC	SPECIES=S.typhimurium; STRAIN=SR-11;			
RX	MEDLINE=98117058; PubMed=9457880;			
RA	Romling U., Bian Z., Hammar M., Sierralta W.D., Normark S.;			
RT	"Curli fibers are highly conserved between Salmonella typhimurium and			
RT	Escherichia coli with respect to operon structure and regulation."			
EL	J. Bacteriol. 180:722-731(1998).			
RN	[2]			
RP	SEQUENCE FROM N.A.			
RC	SPECIES=S.typhimurium; STRAIN=LT2 / SGSC1412 / ATCC 700720;			
RX	MEDLINE=21534948; PubMed=11677609;			
RA	McClelland M., Sanderson K.E., Spieth J., Clifton S.W., Latreille P.,			
RA	Courtney L., Porwollik S., Ali J., Dante M., Du F., Hou S., Layman D.,			
RA	Leonard S., Nguyen C., Scott K., Holmes A., Grewal N., Mulvaney E.,			
RA	Ryan E., Sun A., Florea L., Miller W., Stoneking T., Nhan M.,			
RA	Waterston K., Wilson R.K.;			
RT	"Complete genome sequence of Salmonella enterica serovar Typhimurium			
RT	LT2."			
RL	Nature 413:852-856(2001).			
RN	[3]			
RP	SEQUENCE FROM N.A.			
RC	SPECIES=S.typhi; STRAIN=CT18;			
RX	MEDLINE=21534947; PubMed=11677608;			
RA	Parkhill J., Dougan G., James K.D., Thomson N.R., Pickard D., Wain J.,			
RA	Churcher C., Mungall K.I., Bentley S.D., Holden M.T.G., Sebaihia M.,			
RA	Baker S., Basham D., Brooks K., Chillingworth T., Connor P.,			
RA	Cronin A., Davis P., Davies R.M., Dowd L., White N., Farrar J.,			
RA	Feltwell T., Hamlin N., Haque A., Hien T.T., Holroyd S., Jagels K.,			
RA	Krogh A., Larsen T.S., Leather S., Moule S., O'Gaora P., Parry C.,			
RA	Quail M.A., Rutherford K., Simmonds M., Skelton J., Stevens K.,			
RA	Whitehead S., Barrrell B.G.;			
RT	"Complete genome sequence of a multiple drug resistant Salmonella			
RT	enterica serovar Typhi CT18."			
RL	Nature 413:848-852(2001).			
RN	[4]			
RP	SEQUENCE FROM N.A.			
RC	SPECIES=S.typhi; STRAIN=Ty2 / ATCC 700931;			
RX	MEDLINE=22531367; PubMed=12644504;			
RA	Deng W., Licu S.-R., Plunkett G. III, Mayhew G.F., Rose D.J.,			
RA	Burland V., Kodoyanni V., Schwartz D.C., Blattner F.R.;			

P91698 drosophila  
Q810f0 homo sapien  
P18127 xanthomias  
P30688 neisseria m  
Q24180 drosophila  
Q04893 saccharomyc  
P14291 caenorhabdi  
P34308 caenorhabdi  
P52302 drosophila  
Q12756 homo sapien  
P06997 escherichia  
P55004 pharbitis n

"Comparative genomics of Salmonella enterica serovar Typhi strains Ty2 and CT18";  
J. Bacteriol. 185:2330-2337(2003).  
[5]  
SEQUENCE FROM N.A.  
SPCIES=S.enteritidis; STRAIN=27655-3B;  
MEDLINE=96146512; PubMed=950497;  
RA Collinson S.K., Clouthier S.C., Doran J.L., Baner P.A., Kay W.W.;  
RT "Salmonella enteritidis agfBAC operon encoding thin, aggregative  
fimbriae";  
J. Bacteriol. 178:662-667(1996).  
[6]  
SEQUENCE OF 21-151 FROM N.A.  
RP SPCIES=S.enteritidis; STRAIN=27655-3B;  
RX MEDLINE=94013373; PubMed=8104955;  
RA Duran J.L., Collinson S.K., Burian J., Santos G., Todd E.C.D.,  
RA Munro C.K., Kay C.M., Baner P.A., Peterkin P.I., Kay W.W.;  
RT "DNA-based diagnostic tests for Salmonella species targeting agfA,  
the structural gene for thin, aggregative fimbriae";  
J. Clin. Microbiol. 31:2263-2273(1993).  
[7]  
SEQUENCE OF 21-33.  
RP SPCIES=S.enteritidis; STRAIN=27655-3B;  
RX MEDLINE=91310586; PubMed=1677357;  
RA Collinson S.K., Emeedy L., Mueller K.-M., Trust T.J., Kay W.W.;  
RT "Purification and characterization of thin, aggregative fimbriae from  
Salmonella enteritidis";  
J. Bacteriol. 173:4773-4781(1991).  
CC -!- FUNCTION: CURLIN IS THE STRUCTURAL SUBUNIT OF THE CURLI. CURLI ARE  
COILED SURFACE STRUCTURES THAT ASSEMBLE PREFERENTIALLY AT GROWTH  
TEMPERATURES BELOW 37 DEGREES CELSIUS. CURLI CAN BIND TO  
FIBRONECTIN.  
CC -!- SIMILARITY: BELONGS TO THE CSGA/CSGB FAMILY.  
CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration  
between the Swiss Institute of Bioinformatics and the EMBL Outstation -  
the European Bioinformatics Institute. There are no restrictions on its  
use by non-profit institutions as long as its content is in no way  
modified and this statement is not removed. Usage by and for commercial  
entities requires a license agreement (See http://www.isb-sib.ch/announce/  
or send an email to license@isb-sib.ch).  
EMBL; AJ002301; CAAG5317.1; -  
DR EMBL; AE008749; AAL20074.1; -  
DR EMBL; ALG62769; CAD08268.1; -  
DR EMBL; AE016840; AAC069399.1; -  
DR EMBL; U43280; AAC43599.1; -  
DR FRC; JC6039; JC6039.  
DR SyGene; SG10608; csGA.  
KW Fimbria; Signal; Complete proteome.  
FT SIGNAL 1 20  
FT CHAIN 21 151 MAJOR CURLIN SUBUNIT.  
FT CONFLICT 134 151 SVMVQYGFNGNATANY -> DSYTQVAS (IN  
REP. 6).  
FT SEQUENCE 151 AA; 1305 MW; 87DACD16B21359 CRC64;  
Query Match 89.4%; Score 693; DB 1; Length 151;  
Best Local Similarity 90.7%; Pred. No. 4.Se-53;  
Matches 137; Conservative 3; Mismatches 11; Indels 0; Gaps 0;  
Qy 1 MKLLKVAFAAIYVSGSALGVVPQWGGGNHNGSSGPDSTLSIYQYGSANAAALQ 60  
Db 1 MKLLKVAFAAIYVSGSALGVVPQWGGGNHNGSSGPDSTLSIYQYGSANAAALQ 60  
Qy 61 SDARKSETTTQSQYNGADVGQADNSTIELQTNGFRNNATIDQWNKNKSDITVGYDQ 120  
Db 61 SDARKSETTTQSQYNGADVGQADNSTIELQTNGFRNNATIDQWNKNKSDITVGYDQ 120  
Qy 121 LVTRVTVMHAASVMVQYGFNGNATANY 151  
Db 121 NNAAVLNQTSADSSVMVQYGFNGNATANY 151

RESULT 2  
CSGA\_ECOLI STANDARD; PRT; 151 AA.  
AC P28307;  
DT 01-DEC-1992 (Rel. 24, Created)  
DT 01-OCT-1996 (Rel. 34, Last sequence update)  
DT 28-FEB-2003 (Rel. 41, Last annotation update)  
DE Major curlin subunit precursor.  
GN CSGA OR B1042.  
OS Escherichia coli.  
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;  
OC Enterobacteriaceae; Escherichia.  
OX NCBI\_TaxID=562;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX STRAIN=K12 / W3110;  
RC MEDLINE=93211294; PubMed=8459772;  
RA Olsen A., Arngvist A.;  
RT "The Rpos sigma factor relieves H-NS-mediated transcriptional  
repression of csgA, the subunit gene of fibronectin-binding curli in  
Escherichia coli.";  
RT Mol. Microbiol. 7:523-536(1993).  
[2]  
RP SEQUENCE FROM N.A.  
RX STRAIN=K12 / MC4100;  
RC MEDLINE=96414468; PubMed=8817489;  
RA Hammar M., Arngvist A., Bian Z., Olsen A., Normark S.;  
RT Expression of two csg operons is required for production of  
fibronectin- and Congo red-binding curli polymers in Escherichia coli  
K-12";  
RT Mol. Microbiol. 18:661-670(1995).  
[3]  
RP SEQUENCE FROM N.A.  
RX STRAIN=K12 / MG1655;  
RC MEDLINE=97426617; PubMed=9278503;  
RA Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,  
RA Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,  
RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,  
MAU B., Shao Y.;  
RT The complete genome sequence of Escherichia coli K-12";  
RT Science 277:1233-1238(1997).  
[4]  
RP SEQUENCE FROM N.A.  
RX STRAIN=K12;  
RC MEDLINE=97061202; PubMed=8905232;  
RA Oshima T., Aiba H., Baba T., Fujita K., Hayashi K., Honjo A.,  
RA Ikemoto K., Inada T., Itoh T., Kajihara M., Kanai K., Kashimoto K.,  
RA Kimura S., Kitagawa M., Makino K., Masuda S., Miki T., Mizobuchi K.,  
RA Mori H., Motomura K., Nakamura Y., Nishio Y., Saito N.,  
RA Sampei G., Seki Y., Tagami H., Takemoto K., Wada C., Yamamoto Y.,  
Yano M., Horiuchi T.;  
RT "A 718-kb DNA sequence of the Escherichia coli K-12 genome  
corresponding to the 12.7-28.0 min region on the linkage map";  
RT DNA Res. 3:137-155(1996).  
[5]  
RP SEQUENCE OF 21-40.  
RX STRAIN=K12 / YVEL;  
RC MEDLINE=93023873; PubMed=1357528;  
RA Arngvist A., Olsen A., Pfeiffer J., Russell D.G., Normark S.;  
RT "The Cur protein activates cryptic genes for curli formation and  
fibronectin binding in Escherichia coli HB101";  
RT Mol. Microbiol. 6:2443-2452(1992).  
[6]  
RP SEQUENCE OF 21-31.  
RX MEDLINE=91310586; PubMed=1677357;  
RA Collinson S.K., Emeedy L., Trust T.J., Kay W.W.;  
RT "Purification and characterization of thin, aggregative fimbriae from  
Salmonella enteritidis";  
J. Bacteriol. 173:4773-4781(1991).  
CC -!- FUNCTION: CURLIN IS THE STRUCTURAL SUBUNIT OF THE CURLI. CURLI ARE  
COILED SURFACE STRUCTURES THAT ASSEMBLE PREFERENTIALLY AT GROWTH  
TEMPERATURES BELOW 37 DEGREES CELSIUS. CURLI CAN BIND TO  
FIBRONECTIN.

```

CC -|- SIMILARITY: BELONGS TO THE CSGA/CSGB FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; L04979; AAA23616.1; -
CC EMBL; X90754; CAA62282.1; -
CC EMBL; AB00205; AAC74126.1; -
CC EMBL; D90741; BAA35832.1; -
CC EMBL; D90742; BAA35840.1; -
CC EMBL; D90788; S70788.
CC EMBL; EGI1489; CSGA.
CC EMBL; Signal; Complete proteome.
KW Fimbria; Signal; Complete proteome.
FT SIGNAL 1 20 MAJOR CURLIN SUBUNIT.
FT CHAIN 21 151 A -> E (IN REF. 1).
FT CONFLICT 7 7
FT SEQUENCE 151 AA; 15049 MW; C003470D208D395F CRC64;
SQ
Query Match 68.1%; Score 528; DB 1; Length 151;
Best Local Similarity 68.9%; Pred. No. 7.9e-39; Indels 0; Gaps 0;
Matches 104; Conservative 20; Mismatches 27;
QY 1 MKLLKVAFAAIVVSGSALAGVVPQWGGGNGHGGNSGPDSTLSIYQYGSANAALALQ 60
DB 1 MKLLKVAFAAIVVSGSALAGVVPQYGGGNGHGGNSGPNSELNIYQYGGNSALALQ 60
QY 61 SDARKSETTITQSGYNGADVGAGDANSITELTQNGFRNNAIDQWNAKNSDITVQYDQ 120
DB 61 TDARNSDLTITQHGCGNGADVGQGSDDSIDLTQRFQGSATLDQWNGKNSMTVRQFGG 120
QY 121 LVTRVVTHEMAHASVMVQYVGFQGNATANCY 151
DB 121 GNGAAVDQTASNSTVNTVQVGFQGNATANCY 151
RESULT 3
CSGA_ECOLI57 STANDARD; PRT; 152 AA.
AC Q30U24; 2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Major curlin subunit precursor.
GN CSGA OR Z1676 OR ECS1420.
OS Escherichia coli O157:H7.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Escherichia.
OX NCBI_TaxID=83334;
RN [1]
SEQUENCE FROM N.A.
RC STRAIN=O157:H7 / ATCC 43895;
RX MEDLINE=21218556; PubMed=11319125;
RA Ulrich G.A., Keen J.E., Elder R.O.,
RT "Mutations in the csgD promoter associated with variations in curli
RT expression in certain strains of Escherichia coli O157:H7."
RT Appl. Environ. Microbiol. 67:2367-2370 (2001).
RN [2]
SEQUENCE FROM N.A.
RC STRAIN=O157:H7 / EDL933 / ATCC 700927;
RX MEDLINE=21074935; PubMed=11206551;
RA Perna N.T., Plunkett G. III, Burland V., Mau B., Glasner J.D.,
RA Rose D.J., Mayhew G.F., Evans P.S., Gregor J., Kirkpatrick H.A.,
RA Posfai G., Hackett J., Klink S., Boutin A., Shao Y., Miller L.,
RA Grobeck E.J., Davis N.W., Lim A., Dimalanta E.T., Potamousis K.,
RA Apodaca J., Anantharaman T.S., Lin J., Yen G., Schwartz D.C.,
RA Welch R.A., Blattner F.R.;
RT "Genome sequence of enterohaemorrhagic Escherichia coli O157:H7."
RT Nature 409:529-533 (2001).

```

```

RN SEQUENCE FROM N.A.
RP STRAIN=O157:H7 / RIMD 0509952;
RX MEDLINE=21156231; PubMed=11258796;
RA Hayashi T., Makino K., Ohnishi M., Kurokawa K., Ishii K., Yokoyama K.,
RA Han C.-G., Ohtsubo E., Nakayama K., Murata T., Tanaka M., Tobe T.,
RA Iida T., Takami H., Honda T., Sasakawa C., Ogasawara N., Yasunaga T.,
RA Kuhara S., Shiba T., Hattori M., Shingawa H.;
RT "Complete genome sequence of enterohaemorrhagic Escherichia coli
RT O157:H7 and genomic comparison with a laboratory strain K-12."
RT DNA Res. 8:11-22 (2001).
CC -|- FUNCTION: CURLIN IS THE STRUCTURAL SUBUNIT OF THE CURLI. CURLI ARE
CC COILED SURFACE STRUCTURES THAT ASSEMBLE PREFERENTIALLY AT GROWTH
CC TEMPERATURES BELOW 37 DEGREES CELSIUS. CURLI CAN BIND TO
CC FIBRONECTIN.
CC -|- SIMILARITY: BELONGS TO THE CSGA/CSGB FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; AF275733; AAK53212.1; -
CC EMBL; AB005315; AAG55788.1; -
CC EMBL; AF002554; BAB34843.1; -
CC EMBL; D90806; D90806.
CC EMBL; H85665; H85665.
KW Fimbria; Signal; Complete proteome.
FT SIGNAL 1 20 BY SIMILARITY.
FT CHAIN 21 152 MAJOR CURLIN SUBUNIT.
FT SEQUENCE 152 AA; 15099 MW; EE2D2D94DDE91243 CRC64;
SQ
Query Match 65.4%; Score 506.5; DB 1; Length 152;
Best Local Similarity 67.1%; Pred. No. 5.6e-37; Indels 1; Gaps 1;
Matches 102; Conservative 21; Mismatches 28;
QY 1 MKLLKVAFAAIVVSGSALAGVVPQW-CGGGNGHGGNSGPDSTLSIYQYGSANAALAL 59
DB 1 MKLLKVAFAAIVVSGSALAGVVPQYGGGNGHGGNSGPNSELNIYQYGGNSALAL 60
QY 60 QSDARKSETTITQSGYNGADVGAGDANSITELTQNGFRNNAIDQWNAKNSDITVQYD 119
DB 61 QADARNSDLTITQHGCGNGADVGQGSDDSIDLTQRFQGSATLDQWNGKNSMTVRQFG 120
QY 120 QLVTRVVTHEMAHASVMVQYVGFQGNATANCY 151
DB 121 GNGAAVDQTASNSTVNTVQVGFQGNATANCY 152
RESULT 4
CSGB_ECOLI STANDARD; PRT; 151 AA.
AC P39828;
DT 01-FEB-1995 (Rel. 31, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Minor curlin subunit precursor.
GN CSGB OR B1041 OR Z1675 OR ECS1419.
OS Escherichia coli, and
OS Escherichia coli O157:H7.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Escherichia.
OX NCBI_TaxID=562, 83334;
RN [1]
SEQUENCE FROM N.A.
RC STRAIN=X12 / MC4100;
RX MEDLINE=96414468; PubMed=8817489;
RA Hammar M., Arqvist A., Bian Z., Olsen A., Normark S.;
RT "Expression of two csg operons is required for production of
RT fibronectin- and congo red-binding curli polymers in Escherichia coli

```

```

RT K-12."
RL Mol. Microbiol. 18:661-670(1995).
[2]
RC SEQUENCE FROM N.A.
RX STRAIN=K12 / MG1655;
RC MEDLINE=97426617; PubMed=9278503;
RX Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,
RA Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,
RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,
RA Mau B., Shao Y.;
RT "The complete genome sequence of Escherichia coli K-12."
RL Science 277:1453-1474(1997).
[3]
RC SEQUENCE FROM N.A.
RX STRAIN=K12;
RC MEDLINE=97061202; PubMed=8905232;
RA Oshima T., Aiba H., Baba T., Fujita K., Hayashi K., Honjo A.,
RA Ikemoto K., Inada T., Itoh T., Kajihara M., Kanai K., Kashimoto K.,
RA Kimura S., Kitagawa M., Makino K., Masuda S., Miki T., Mizobuchi K.,
RA Mori H., Motomura K., Nakamura Y., Nishimoto H., Nishio Y., Saito N.,
RA Sampei G., Seki Y., Tagami H., Takemoto K., Wada C., Yamamoto Y.,
RA Yano M., Horiuchi T.;
RT "A 718-kb DNA sequence of the Escherichia coli K-12 genome
RT corresponding to the 12.7-28.0 min region on the linkage map."
RL DNA Res. 3:137-155(1996).
[4]
RC SEQUENCE FROM N.A.
RX STRAIN=O157:H7 / EDL933 / ATCC 700927;
RC MEDLINE=21074935; PubMed=11206551;
RA Perna N.T., Plunkett G. III, Burland V., Mau B., Glasner J.D.,
RA Rose D.J., Mayhew G.F., Evans P.S., Gregor J., Kirkpatrick H.A.,
RA Posfai J., Hackett J., Klink S., Boutin A., Shao Y., Miller L.,
RA Grobeck E.J., Davis N.W., Lim A., Dimalanta E.T., Potamouisis K.,
RA Apodaca J., Anantharaman T.S., Lin J., Yen G., Schwartz D.C.,
RA Weich R.A., Blattner F.R.;
RT "Genome sequence of enterohaemorrhagic Escherichia coli O157:H7."
RL Nature 409:529-533(2001).
[5]
RC SEQUENCE FROM N.A.
RX STRAIN=O157:H7 / RIMD 0509952;
RC MEDLINE=21156231; PubMed=11258796;
RA Hayashi T., Makino K., Onishi M., Kurokawa K., Ishii K., Yokoyama K.,
RA Han C.-G., Ohtsubo E., Nakayama K., Murata T., Tanaka N., Tohe T.,
RA Iida T., Takami H., Honda T., Sasakawa C., Ogasawara N., Yasunaga T.,
RA Kuhara S., Shiba T., Hattori M., Shinagawa H.;
RT "Complete genome sequence of enterohemorrhagic Escherichia coli
RT O157:H7 and genomic comparison with a laboratory strain K-12."
RL DNA Res. 8:11-22(2001).
[6]
RC SEQUENCE OF 1-21 FROM N.A.
RX STRAIN=K12;
RC MEDLINE=95157246; PubMed=7854117;
RA Arngvist A., Olsen A., Normark S.;
RT "Sigma S-dependent growth-phase induction of the csqBA promoter in
RT Escherichia coli can be achieved in vivo by sigma 70 in the absence
RT of the nucleoid-associated protein H-NS."
RL Mol. Microbiol. 13:1021-1032(1994).
CC -!- FUNCTION: CURLIN IS THE STRUCTURAL SUBUNIT OF THE CURLI. CURLI ARE
CC COILED SURFACE STRUCTURES THAT ASSEMBLE PREFERENTIALLY AT GROWTH
CC TEMPERATURES BELOW 37 DEGREES CELSIUS. CURLI CAN BIND TO
CC FIBRONECTIN. THE MINOR SUBUNIT IS THE NUCLEATION COMPONENT OF
CC CURLIN MONOMERS.
CC -!- SIMILARITY: BELONGS TO THE CSGA/CSGB FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; X90754; CAA62281.1; -

```

```

DR EMBL; AE000205; AAC74125.1; -.
DR EMBL; D90741; BAA35831.1; -.
DR EMBL; AE005315; AAG55787.1; -.
DR EMBL; AP002554; BAB34842.1; -.
DR PIR; C90806; C90806.
DR PIR; G85665; G85665.
DR PIR; S70787; S70787.
DR EcoGene; EGI2621; csqB.
KW Fibria; Signal; Complete proteome.
FT SIGNAL 21 POTENTIAL.
FT CHAIN 22 151 MINOR CURLIN SUBUNIT.
SQ SEQUENCE 151 AA; 15882 MW; B18D266B964014B8 CRC64;

Query Match 13.5%; Score 104.5; DB 1; Length 151;
Best Local Similarity 29.2%; Pred. No. 0.023;
Matches 35; Conservative 19; Mismatches 59; Indels 7; Gaps 4;

QY 9 FAATVSGSAL--AGVVPQWGGGNGHNGGNSGPDSTLSIY-QYGSANAALALQSDARK 65
Db 33 FAVNELSKSFNQAAII---QAGTNNSAOLROGSKLLAVVAQESSNPA-KIDQTGDY 88
QY 66 SETTITSGYGVGADVGQGDADNSTIETQTQFRNNTIDQWNAKNSDITVGGYDQLVTRV 125
Db 89 NLAYIDQAGSANDASISQAGVAGTAMTIQKSGNKANITQYGTQKTAIVVQROSQMAIRV 148

RESULT 5
CSGB_SALTY STANDARD; PRT; 151 AA.
AC P55226;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Minor curlin subunit precursor (Fimbriae SEF17 minor subunit).
GN CSGB OR AGFB OR STM143.
OS Salmonella typhimurium, and
OS Salmonella enteritidis.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Salmonella.
OX NCBI_taxID=602, 592;
RX [1]
SEQUENCE FROM N.A.
RP SPECIES=S.typhimurium; STRAIN=SR-11;
RC MEDLINE=98117058; PubMed=9457880;
RA Romling U., Bian Z., Hammar M., Sieralta W.D., Normark S.;
RT "Curli fibers are highly conserved between Salmonella typhimurium and
RT Escherichia coli with respect to operon structure and regulation."
RL J. Bacteriol. 180:722-731(1998).
[2]
SEQUENCE FROM N.A.
RP SPECIES=S.typhimurium; STRAIN=L72 / SGSC1412 / ATCC 700720;
RC MEDLINE=21534948; PubMed=11677609;
RA McClelland M., Sanderson K.E., Spieth J., Clifton S.W., Latreille P.,
RA Courtney L., Porwollik S., Ali J., Dante M., Du P., Hou S., Layman D.,
RA Leonard S., Nguyen C., Scott K., Holmes A., Grewal N., Mulvaney E.,
RA Ryan E., Sun H., Florea L., Miller W., Stoneking T., Nhan M.,
RA Waterston R., Wilson R.K.;
RT "Complete genome sequence of Salmonella enterica serovar Typhimurium
RT LT2."
RL Nature 413:852-856(2001).
[3]
SEQUENCE FROM N.A.
RP SPECIES=S.enteritidis; STRAIN=27655-3B;
RC MEDLINE=9614651; PubMed=8550497;
RA Collinson S.K., Clouthier S.C., Doran J.L., Baner P.A., Kay W.W.;
RT "Salmonella enteritidis agfBAC operon encoding thin, aggregative
RT fimbriae."
RL J. Bacteriol. 178:662-667(1996).
CC -!- FUNCTION: CURLIN IS THE STRUCTURAL SUBUNIT OF THE CURLI. CURLI ARE
CC COILED SURFACE STRUCTURES THAT ASSEMBLE PREFERENTIALLY AT GROWTH
CC TEMPERATURES BELOW 37 DEGREES CELSIUS. CURLI CAN BIND TO
CC FIBRONECTIN. THE MINOR SUBUNIT IS THE NUCLEATION COMPONENT OF
CC CURLIN MONOMERS.

```



```

CC -1- SIMILARITY: BELONGS TO THE CSGA/CSGB FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: AJ002301; CAA05316.1; -
CC EMBL: AB008749; AAL20073.1; -
CC EMBL: U43280; AAC33598.1; -
CC PIR: JC6040; JC6040.
CC StyGene; SGI0609; csGb.
CC FimBria; Signal; Complete proteome.
CC SIGNAL 1 21 POTENTIAL.
CC CHAIN 22 151 MINOR CURLIN SUBUNIT.
CC SEQUENCE 151 AA; 16182 MW; 60FC5430E6DD361D CRC64;
CC -----
CC Query Match 12.7%; Score 98.5; DB 1; Length 151;
CC Best Local Similarity 26.2%; Pred. No. 0.074;
CC Matches 34; Conservative 17; Mismatches 42; Indels 37; Gaps 4;
CC -----
CC QY 10 AAIIVSGSALAGVVPWGGGNGGNGSGDPSTLSIYQGSANAALALQSDARKSETT 69
CC Db 58 ARVROEGSKLLSVISQ--EGGNRAKVDQAGNYNFAYIEOTGNAN-----DAS 103
CC QY 70 ITQSGVNGADVQGGADNSTIETQNGFRNATIDOWNAKNSDITVGYDQLVTRVYTHE 129
CC Db 104 ISSAIVGNSA-----AIIQSGNKANITQYGTCK-----TAVVQK 140
CC QY 130 MAHASVWVRQ 139
CC Db 141 QSHWAIRVTQ 150
CC -----
CC RESULT 6
CC TNKI_HUMAN STANDARD; PRT; 1327 AA.
CC ID TNKI_HUMAN
CC AC Q95271; Q95272;
CC DT 28-FEB-2003 (Rel. 41, Created)
CC DT 28-FEB-2003 (Rel. 41, Last sequence update)
CC DT 10-OCT-2003 (Rel. 42, Last annotation update)
CC DE Tankyrase 1 (EC 2.4.2.30) (TNKI) (Tankyrase I) (TNKS-1) (TRF1-
CC interacting ankyrin-related ADP-ribose polymerase).
CC GN TNKS OR TNKS1 OR TIN1 OR TIN1 OR PARPL.
CC OS Homo sapiens (Human).
CC OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
CC OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
CC OX NCBI_TaxID=9606;
CC [1]
CC SEQUENCE FROM N.A. (ISOFORMS 1 AND 2).
CC RC TISSUE=Testis;
CC RA Smith S., Giriat I., Schmitt A., de Lange T.;
CC RT "Tankyrase, a poly(ADP-ribose) polymerase at human telomeres.";
CC RL Science 282:1484-1487 (1998).
CC [2]
CC SUBCELLULAR LOCATION.
CC RP MEDLINE=99454782; PubMed=10523501;
CC RA Smith S., de Lange T.;
CC RT "Cell cycle dependent localization of the telomeric PARP, tankyrase,
CC to nuclear pore complexes and centrosomes.";
CC RL J. Cell Sci. 112:3649-3656 (1999).
CC [3]
CC FUNCTION, AND PHOSPHORYLATION.
CC RP MEDLINE=20556282; PubMed=10988299;
CC RA Chi N.-W., Lodish H.F.;
CC RT "Tankyrase is a Golgi-associated mitogen-activated protein kinase
CC substrate that interacts with IRAP in GLUT4 vesicles.";
CC RL J. Biol. Chem. 275:38437-38444 (2000).
CC [4]

```

```

RP FUNCTION, AND MUTAGENESIS OF HIS-1184 AND GLU-1291.
RX MEDLINE=21602874; PubMed=11739745;
RA Cook B.D., Dynek J.N., Chang W., Shostak G., Smith S.;
RT "Role for the related poly(ADP-Ribose) polymerases tankyrase 1 and 2
RT at human telomeres.";
RL Mol. Cell Biol. 22:332-342 (2002).
CC -1- FUNCTION: May regulate vesicle trafficking and modulate the
CC subcellular distribution of SLC2A4/GLUT4-vesicles. Has PARP
CC activity and can modify TRF1, and thereby contribute to the
CC regulation of telomere length.
CC -1- CATALYTIC ACTIVITY: NAD(+) + {ADP-D-ribose}(N)-acceptor =
CC nicotinamide + {ADP-D-ribose}(N+1)-acceptor.
CC -1- SUBUNIT: Oligomerizes and associates with TNKS2. Interacts with
CC the cytoplasmic domain of LIMEP/Otase in SLC2A4/GLUT4-vesicles.
CC Binds to the N-terminus of telomeric TRF1 via the ANK repeats.
CC -1- SUBCELLULAR LOCATION: Cytoplasmic, associated with the Golgi and
CC also found at nuclear pore complexes and around the pericentriolar
CC matrix of mitotic centrosomes. During interphase, a small fraction
CC of TNKS is found in the nucleus, associated with TRF1.
CC -1- ALTERNATIVE PRODUCTS:
CC Event=Alternative splicing; Named isoforms=2;
CC Name=1;
CC IsoId=Q95271-1; Sequence=Displayed;
CC Name=2;
CC IsoId=Q95271-2; Sequence=VSP_004538, VSP_004539;
CC Note=No experimental confirmation available;
CC -1- TISSUE SPECIFICITY: Ubiquitous; highest levels in testis.
CC -1- PTM: Upon insulin-stimulation, phosphorylated on serine residues
CC by MAPK kinases.
CC -1- PTM: ADP-ribosylated (-auto).
CC -1- SIMILARITY: Belongs to the PARP family.
CC -1- SIMILARITY: Contains 15 ANK repeats.
CC -1- SIMILARITY: Contains 1 sterile alpha motif (SAM) domain.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: AF082556; AAC79841.1; -
CC EMBL: AF082557; AAC79842.1; -
CC EMBL: AF082558; AAC79843.1; -
CC EMBL: AF082559; AAC79844.1; -
CC HSP: Q00420; IAWC.
CC Genew; HGNC:11941; TNKS.
CC MIM; 603303; -
CC GO; GO:0000781; C:chromosome, telomeric region; IDA.
CC GO; GO:0003950; F:NAD ADP-ribosyltransferase activity; IDA.
CC GO; GO:0005515; P:protein binding; IPI.
CC GO; GO:0007004; P:telomerase-dependent telomere maintenance; NAS.
CC InterPro; IPR002110; ANK.
CC InterPro; IPR001660; SAM.
CC Pfam; PF00023; ank; 19.
CC Pfam; PF00536; SAM; 1.
CC PRINTS; PR01415; ANKYRIN.
CC SMART; SM00248; ANK; 17.
CC SMART; SM00454; SAM; 1.
CC PROSITE; PS00688; ANK_REPEAT; 15.
CC PROSITE; PS00997; ANK_REPEAT_REGION; 1.
CC PROSITE; PS01005; SAM_DOMAIN; 1.
CC Transferase; Glycosyltransferase; NAD; Golgi stack; Telomere;
CC Nuclear protein; Repeat; ANK repeat; ADP-ribosylation;
CC Phosphorylation; Alternative splicing.
CC REPEAT 215 247 ANK 1.
CC REPEAT 248 280 ANK 2.
CC REPEAT 281 333 ANK 3.
CC REPEAT 368 400 ANK 4.
CC REPEAT 401 433 ANK 5.
CC REPEAT 434 466 ANK 6.

```



```

DT 28-FEB-2003 (Rel. 41, Last sequence update)
DE 10-OCT-2003 (Rel. 42, Last annotation update)
DE Minor curlin subunit precursor.
GN CSGB OR STY1180 OR TI1777.
OS Salmonella typhi.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Salmonella.
OX NCBI_TaxID=601;
RN [1]
RN SEQUENCE FROM N.A.
RC STRAIN=CT18;
RC MEDLINE=21534947; PubMed=11677608;
RA Parkhill J., Dougan G., James K.D., Thomson N.R., Pickard D., Wain J.,
RA Churcher C., Mungall K.L., Bentley S.D., Holden M.T.G., Sebahia M.,
RA Baker S., Basham D., Brooks K., Chillingworth T., Connor P.,
RA Cronin A., Davis P., Davies R.M., Dowd L., White N., Farrar J.,
RA Feltwell T., Hamlin N., Haque A., Hien T.T., Holroyd S., Jagels K.,
RA Krogh A., Larsen T.S., Leather S., Moule S., O'Gaora P., Parry C.,
RA Quail M.A., Rutherford K., Simmonds M., Skelton J., Stevens K.,
RA Whitehead S., Barrrell B.G.;
RT "Complete genome sequence of a multiple drug resistant Salmonella
RT enterica serovar Typhi CT18."
RL Nature 413:848-852(2001).
RN [2]
RN SEQUENCE FROM N.A.
RC STRAIN=Ty2 / ATCC 700931;
RC MEDLINE=22531367; PubMed=12644504;
RA Deng W., Liou S.-R., Plunkett G. III, Mayhew G.F., Rose D.J.,
RA Burland V., Kodyolanni V., Schwartz D.C., Blattner P.R.;
RT "Comparative genomics of Salmonella enterica serovar Typhi strains Ty2
RT and CT18."
RL J. Bacteriol. 185:2330-2337(2003).
CC -!- FUNCTION: CURLIN IS THE STRUCTURAL SUBUNIT OF THE CURLI. CURLI ARE
CC COILED SURFACE STRUCTURES THAT ASSEMBLE PREFERENTIALLY AT GROWTH
CC TEMPERATURES BELOW 37 DEGREES CELSIUS. CURLI CAN BIND TO
CC FIBRONECTIN. THE MINOR SUBUNIT IS THE NUCLEATION COMPONENT OF
CC CURLIN MONOMERS.
CC -!- SIMILARITY: BELONGS TO THE CSGB/CSGB FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; AL627269; CAD08267.1; -
CC DR EMBL; AB016840; AAO69400.1; -
CC KW Fimbrin; Signal; Complete proteome.
CC FT SIGNAL 1 21 POTENTIAL.
CC FT CHAIN 22 151 MINOR CURLIN SUBUNIT.
CC SQ SEQUENCE 151 AA; 16254 MW; 161C54326E573495 CRC64;
Query Match 11.78; Score 90.5; DB 1; Length 151;
Best Local Similarity 25.4%; Pred. No. 0.36; Indels 37; Gaps 4;
Matches 33; Conservative 17; Mismatches 43;
QY 10 AAIWVGSALAGVVPWGGGNGHNGGNSGSPDSTLSIYQYGSANAALALQSDARKSETT 69
Db 58 ARVREQSGSKLLSVISQ--EGENNRAKYDQAGNYNFAYIEQTGNAN-----DAS 103
QY 70 ITQSGYNGADVGGGADNSITELTQNGFRNNAIDQNAKNSDITVGYQDLYTRVVTHE 129
Db 104 ISQAYGNSA-----AIIQKSGNKNITQYGTOK-----TAVVVQK 140
QY 130 MAHASVMVQK 139
Db 141 QSHMAIRVQK 150
RESULT 9
PER_DROWI

```

```

ID PER DROWI STANDARD; PRT; 1093 AA.
AC Q03297; O18421; O18422; P91721; P91722;
DT 01-OCT-1993 (Rel. 27, Created)
DT 15-JUL-1998 (Rel. 36, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Period circadian protein (Fragment).
DE PER.
GN Drosophila willistoni (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7260;
RN [1]
RN SEQUENCE FROM N.A.
RC STRAIN=Various strains;
RC MEDLINE=97357421; PubMed=9214747;
RA Gleason J.M., Powell J.R.;
RA "Intraspecific and intraspecific comparisons of the period locus in
RA the Drosophila willistoni sibling species."
RL Mol. Biol. Evol. 14:741-753(1997).
RN [2]
RN SEQUENCE OF 579-646 FROM N.A.
RC MEDLINE=93196482; PubMed=8450754;
RA Peixoto A.A., Campegan S., Costa R.H., Kyriacou C.P.;
RT "Molecular evolution of a repetitive region within the per gene of
RT Drosophila."
RL Mol. Biol. Evol. 10:127-139(1993).
CC -!- FUNCTION: Essential for biological clock functions. Determines the
CC period length of circadian and ultradian rhythms; an increase in
CC PER dosage leads to shortened circadian rhythms and a decrease
CC leads to lengthened circadian rhythms. Essential for the circadian
CC rhythmicity of locomotor activity, eclosion behavior, and for the
CC rhythmic component of the male courtship song that originates in
CC the thoracic nervous system. The biological cycle depends on the
CC rhythmic formation and nuclear localization of the TIM-PER
CC complex. Light induces the degradation of TIM, which promotes
CC elimination of PER. Nuclear activity of the heterodimer
CC coordinatively regulates PER and TIM transcription through a
CC negative feedback loop. Behaves as a negative element in circadian
CC transcriptional loop. Does not appear to bind DNA, suggesting
CC indirect transcriptional inhibition (By similarity).
CC -!- SUBUNIT: Forms heterodimer with timeless (TIM); the complex then
CC translocates into the nucleus (By similarity).
CC -!- SUBCELLULAR LOCATION: Nuclear at specific periods of the day.
CC First accumulates in the perinuclear region about one hour before
CC translocation into the nucleus. Interaction with Tim is required
CC for nuclear localization (By similarity).
CC -!- PTM: PHOSPHORYLATED WITH A CIRCADIAN RHYTHMICITY, PROBABLY BY THE
CC DOUBLE-TIME PROTEIN (DET). PHOSPHORYLATION COULD BE IMPLICATED IN
CC THE STABILITY OF PER MONOMER AND IN THE FORMATION OF HETERODIMER
CC PER-TIM (By similarity).
CC -!- MISCELLANEOUS: THE SEQUENCE SHOWN IS THAT OF STRAIN ATLIXCO.
CC -!- SIMILARITY: Contains 2 PAS (PER-ARNT-SIM) dimerization domains.
CC -!- SIMILARITY: Contains 1 PAS-associated C-terminal (PAC) domain.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; U51055; AAB41360.1; -
CC DR EMBL; U51056; AAB41361.1; -
CC DR EMBL; U51057; AAB41362.1; -
CC DR EMBL; U51058; AAB41363.1; -
CC DR EMBL; U51059; AAB41364.1; -
CC DR EMBL; U51060; AAB41365.1; -
CC DR EMBL; U51061; AAB41366.1; -
CC DR EMBL; U51062; AAB41367.1; -
CC DR EMBL; U51063; AAB41368.1; -
CC DR EMBL; U51064; AAB41369.1; -

```

DR EMBL; U51065; AAB41370.1; --  
 DR EMBL; U51066; AAB41371.1; --  
 DR EMBL; U51067; AAB41372.1; --  
 DR EMBL; U51068; AAB41373.1; --  
 DR EMBL; U51069; AAB41374.1; --  
 DR EMBL; U51070; AAB41375.1; --  
 DR EMBL; U51071; AAB41376.1; --  
 DR EMBL; U51072; AAB41377.1; --  
 DR EMBL; L06342; AAA28765.1; --  
 DR FlyBase; FBgn0013151; Dm11per.  
 DR InterPro; IPR000014; PAS\_domain.  
 DR Pfam; PF00989; PAS; 2.  
 DR SMART; SMO0091; PAS; 2.  
 DR PROSITE; PS00112; PAS; 2.  
 DR Biological rhythms; Repeat; Nuclear protein; Phosphorylation;  
 KW Polymorphism.  
 FT NON\_TER 1 12 NUCLEAR LOCALIZATION SIGNAL (POTENTIAL).  
 FT DOMAIN <1 12 PAS 1.  
 FT DOMAIN 139 209 PAS 2.  
 FT DOMAIN 289 359 PAS 2.  
 FT DOMAIN 371 411 PAC.  
 FT DOMAIN 7 12 POLY-LYS.  
 FT DOMAIN 618 625 POLY-GLY.  
 FT DOMAIN 718 734 POLY-GLY.  
 FT DOMAIN 745 778 POLY-SER.  
 FT DOMAIN 759 778 POLY-GLY.  
 FT DOMAIN 885 888 POLY-ALA.  
 FT DOMAIN 911 917 POLY-ALA.  
 FT VARIANT 611 611 T -> A (IN STRAIN MANAUS 2).  
 FT VARIANT 617 617 S -> F (IN STRAIN 0811.4).  
 FT VARIANT 622 622 G -> V (IN STRAIN GUANA).  
 FT VARIANT 724 724 G -> A (IN STRAIN MANAUS 2).  
 FT VARIANT 726 726 G -> S (IN STRAIN SANTA MARIA).  
 FT VARIANT 729 734 MISSING (IN STRAIN PORTO ALEGRE 3).  
 FT VARIANT 730 734 MISSING (IN STRAIN PORTO ALEGRE 4).  
 FT VARIANT 731 734 MISSING (IN STRAINS MANAUS 1 AND MANAUS 3).  
 FT VARIANT 732 734 MISSING (IN STRAINS LIMA B, L'HABITATUE AND CANO MORA).  
 FT VARIANT 733 733 G -> V (IN STRAIN PORTO ALEGRE 4).  
 FT VARIANT 733 734 MISSING (IN STRAINS GUADELOUPE, MANAUS 2, PORTO ALEGRE 2, PORTO ALEGRE 1 AND GUANA).  
 FT VARIANT 734 734 MISSING (IN STRAINS MANAUS 4, PORTO ALEGRE 1 AND PORTO ALEGRE 2).  
 FT VARIANT 747 747 S -> A (IN STRAINS GUADELOUPE AND GUANA).  
 FT VARIANT 764 766 MISSING (IN STRAINS MANAUS 3).  
 FT VARIANT 886 886 A -> T (IN STRAIN 0811.4).  
 FT NON\_TER 1093 1093  
 SQ SEQUENCE 1093 AA; 115896 MW; AB6DE050267EC187 CRC64;  
 Query Match 11.6%; Score 90; DB 1; Length 1093;  
 Best Local Similarity 27.7%; Pred.No. 3.9;  
 Matches 26; Conservative 6; Mismatches 46; Indels 16; Gaps 2;  
 QY 27 GGGGNGHNGGNSGPDSTLSIYQVGSANAALQSDARKSETTITQSGYNGADVGQAD 86  
 Db 721 GGGGGGGGGGGGLFLDLVHTSS-----SSQNKGPVTGAAGGAGGGVGGGG-- 770  
 QY 87 NSTIETQNGFRNNTATIDOWNAKNSDITVGQYDQ 120  
 Db 771 -----SCSGLGNGNGVSGNGNSQFSTNQYQ 798  
 RESULT 10  
 VG38\_BPT2  
 ID VG38\_BPT2 STANDARD; PRT; 262 AA.  
 AC P07875;  
 DT 01-AUG-1988 (Rel. 08, Created)  
 DT 01-AUG-1988 (Rel. 08, Last sequence update)  
 DT 01-AUG-1990 (Rel. 15, Last annotation update)  
 DE Receptor recognizing protein (Protein Gp38).  
 GN 38.

OS Bacteriophage T2.  
 OC Viruses: dsDNA viruses, no RNA stage; Caudovirales; Myoviridae;  
 OC T4-like viruses  
 OX NCBI\_TaxID=10664;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=87283911; PubMed=3302276;  
 RA Riede I., Drexler K., Eschbach M.L., Henning U.;  
 RT "DNA sequence of genes 38 encoding a receptor-recognizing protein of  
 bacteriophages T2, K3 and of K3 host range mutants.";  
 RL J. Mol. Biol. 194:31-39(1987).  
 CC -!- FUNCTION: Vg38 is at the tip of the long tail fibers and serves as  
 the phage recognition site for the cellular receptor.  
 CC -!- MISCELLANEOUS: THIS PHAGE USE OUTER MEMBRANE PROTEINS OMPF AND TTR  
 AS RECEPTORS.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 the European Bioinformatics Institute. There are no restrictions on its  
 use by non-profit institutions as long as its content is in no way  
 modified and this statement is not removed. Usage by and for commercial  
 entities requires a license agreement (see http://www.isb-sib.ch/announce/  
 or send an email to license@isb-sib.ch).  
 CC -----  
 DR EMBL; X05312; CAA28935.1; --  
 DR PIR; S00275; S00275.  
 DR InterPro; IPR007932; Tail\_fibre\_GP38.  
 DR Pfam; PF05268; GP38; 1.  
 KW Fiber protein; Phage recognition.  
 SQ SEQUENCE 262 AA; 25801 MW; 0567366918F6C745 CRC64;  
 Query Match 11.5%; Score 89.5; DB 1; Length 262;  
 Best Local Similarity 34.1%; Pred.No. 0.83;  
 Matches 30; Conservative 8; Mismatches 33; Indels 17; Gaps 4;  
 QY 27 GGGGNGHNGGNSGPDSTLSIYQVGSANAALQSDARKSETTITQSGYNGADVGQAD 86  
 Db 175 GGGGRPFVGGKIGSDSILS-----GSNASL---TDAGTGGTTF-QYGAGNGGVGAGG 225  
 QY 87 NSTIETQNGFRNNTATIDOWNAKNSDIT 114  
 Db 226 -----RWGKNVYTSGGAGAAVT 245  
 RESULT 11  
 ID PRTC\_ERWCH STANDARD; PRT; 479 AA.  
 AC P16317;  
 DT 01-AUG-1990 (Rel. 15, Created)  
 DT 01-NOV-1991 (Rel. 20, Last sequence update)  
 DT 10-OCT-2003 (Rel. 42, Last annotation update)  
 DE Secreted protease C precursor (EC 3.4.24.-) (ProC).  
 GN PRTC.  
 OS Erwinia chrysanthemi.  
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;  
 OC Enterobacteriaceae; Pectobacterium.  
 OX NCBI\_TaxID=556;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX STRAIN=B374;  
 RX MEDLINE=91009140; PubMed=2211614;  
 RA Deleplaire P., Wandersman C.;  
 RT "Protein secretion in Gram-negative bacteria. The extracellular  
 metalloprotease B from Erwinia chrysanthemi contains a C-terminal  
 secretion signal analogous to that of Escherichia coli alpha-  
 hemolysin.";  
 RL J. Biol. Chem. 265:17118-17125(1990).  
 RN [2]  
 RP SEQUENCE OF 1-59 FROM N.A., AND PARTIAL SEQUENCE.  
 RC STRAIN=B374;  
 RX MEDLINE=89255387; PubMed=2722818;  
 RA Deleplaire P., Wandersman C.;  
 RT "Protease secretion by Erwinia chrysanthemi. Proteases B and C are

RT synthesized and secreted as zymogens without a signal peptide.;  
 RL J. Biol. Chem. 264:9083-9089(1989).  
 CC -!- COFACTOR: Binds 1 zinc ion and 7 calcium ions per subunit (By  
 CC similarity).  
 CC -!- SUBCELLULAR LOCATION: Secreted.  
 CC -!- DOMAIN: THE GLY-RICH REPEATS MAY BE IMPORTANT IN THE EXTRACELLULAR  
 CC SECRETION OF THIS METALLOPROTEASE.  
 CC -!- SIMILARITY: Belongs to peptidase family M10B.  
 CC  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See [http://www.isb-sib.ch/](http://www.isb-sib.ch/announce/) or  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC  
 CC EMBL; M59229; AAA24860.1; -;  
 CC EMBL; J04736; AAA24862.1; -;  
 CC EMBL; M60395; AAA63638.1; -;  
 CC PIR; A38307; A38307.  
 CC PDB; 1G07; 17-OCT-02.  
 CC PDB; 1G08; 17-OCT-02.  
 CC PDB; 1K7G; 19-OCT-02.  
 CC PDB; 1K7Q; 20-OCT-02.  
 CC MEROPS; M10.054; -;  
 CC InterPro; IPR001343; Hemolysin Ca bind.  
 CC InterPro; IPR006025; Pept M.Zn.Bs.  
 CC InterPro; IPR006026; Peptidase.M.  
 CC Pfam; PF00553; hemolysincabind; 3.  
 CC PRINTS; PR00313; CABNDNGRPT.  
 CC SMART; SM00235; ZnMC; 1.  
 CC PROSITE; PS00142; ZINC PROTEASE; 1.  
 CC PROSITE; PS00330; HEMOLYSIN CALCIUM; 1.  
 CC Hydrolase; Metalloprotease; Calcium-binding; Metal-binding; Zinc;  
 KW Repeat; Zymogen; 3D-structure.  
 KW  
 FT PROPEP 1 17  
 FT CHAIN 18 479 SECRETED PROTEASE C.  
 FT REPEAT 345 353 GXGXD 1.  
 FT REPEAT 354 362 GXGXD 2.  
 FT REPEAT 363 371 GXGXD 3.  
 FT REPEAT 372 380 GXGXD 4.  
 FT REPEAT 381 389 GXGXD 5.  
 FT METAL 188 188 ZINC (CATALYTIC) (BY SIMILARITY).  
 FT ACT\_SITE 189 189 BY SIMILARITY.  
 FT METAL 192 192 ZINC (CATALYTIC) (BY SIMILARITY).  
 FT METAL 228 228 ZINC (CATALYTIC) (BY SIMILARITY).  
 FT METAL 265 265 CALCIUM 1 (VIA CARBONYL OXYGEN) (BY  
 FT SIMILARITY).  
 FT METAL 267 267 CALCIUM 1 (VIA CARBONYL OXYGEN) (BY  
 FT SIMILARITY).  
 FT METAL 297 297 CALCIUM 1 (BY SIMILARITY).  
 FT METAL 299 299 CALCIUM 1 (VIA CARBONYL OXYGEN) (BY  
 FT SIMILARITY).  
 FT METAL 300 300 CALCIUM 2 (VIA CARBONYL OXYGEN) (BY  
 FT SIMILARITY).  
 FT METAL 302 302 CALCIUM 1 AND 2 (BY SIMILARITY).  
 FT METAL 339 339 CALCIUM 2 (VIA CARBONYL OXYGEN) (BY  
 FT SIMILARITY).  
 FT METAL 341 341 CALCIUM 2 (BY SIMILARITY).  
 FT METAL 346 346 CALCIUM 3 (VIA CARBONYL OXYGEN) (BY  
 FT SIMILARITY).  
 FT METAL 348 348 CALCIUM 3 (VIA CARBONYL OXYGEN) (BY  
 FT SIMILARITY).  
 FT METAL 350 350 CALCIUM 3 (BY SIMILARITY).  
 FT METAL 355 355 CALCIUM 4 (VIA CARBONYL OXYGEN) (BY  
 FT SIMILARITY).  
 FT METAL 357 357 CALCIUM 4 (VIA CARBONYL OXYGEN) (BY  
 FT SIMILARITY).  
 FT METAL 359 359 CALCIUM 4 (BY SIMILARITY).  
 FT METAL 363 363 CALCIUM 3 (VIA CARBONYL OXYGEN) (BY  
 FT SIMILARITY).  
 FT METAL 364 364 CALCIUM 5 (VIA CARBONYL OXYGEN) (BY  
 FT SIMILARITY).

FT METAL 365 365 SIMILARITY).  
 FT CALCIUM 3 (VIA CARBONYL OXYGEN) (BY  
 FT SIMILARITY).  
 FT METAL 366 366 CALCIUM 5 (VIA CARBONYL OXYGEN) (BY  
 FT SIMILARITY).  
 FT METAL 368 368 CALCIUM 3 AND 5 (BY SIMILARITY).  
 FT METAL 372 372 CALCIUM 4 (VIA CARBONYL OXYGEN) (BY  
 FT SIMILARITY).  
 FT METAL 373 373 CALCIUM 6 (VIA CARBONYL OXYGEN) (BY  
 FT SIMILARITY).  
 FT METAL 375 375 CALCIUM 6 (VIA CARBONYL OXYGEN) (BY  
 FT SIMILARITY).  
 FT METAL 377 377 CALCIUM 4 AND 6 (BY SIMILARITY).  
 FT METAL 381 381 CALCIUM 5 (VIA CARBONYL OXYGEN) (BY  
 FT SIMILARITY).  
 FT METAL 382 382 CALCIUM 7 (VIA CARBONYL OXYGEN) (BY  
 FT SIMILARITY).  
 FT METAL 383 383 CALCIUM 5 (VIA CARBONYL OXYGEN) (BY  
 FT SIMILARITY).  
 FT METAL 384 384 CALCIUM 7 (VIA CARBONYL OXYGEN) (BY  
 FT SIMILARITY).  
 FT METAL 386 386 CALCIUM 5 AND 7 (BY SIMILARITY).  
 FT METAL 395 395 CALCIUM 6 (BY SIMILARITY).  
 FT METAL 402 402 CALCIUM 6 (BY SIMILARITY).  
 FT METAL 412 412 CALCIUM 7 (BY SIMILARITY).  
 SQ SEQUENCE 479 AA; 51600 MW; 990ED8376725DF61 CRC64;  
 Query Match 11.4%; Score 88.5; DB 1; Length 479;  
 Best Local Similarity 28.8%; Pred. No. 2;  
 Matches 40; Conservative 21; Mismatches 41; Indels 37; Gaps 9;  
 QY 27 GGGNGHNGSGNSGPDSTLSIYQYGSANAALALQSDARKSETTITQSG----YNGG--- 78  
 DB 345 GSGNDILVNSAD-----NLQGGAGNDVLYGGAGA---DTLYGGAGRTFFVYGGQDS 396  
 QY 79 -----ADYGGQAGNSTIELTONGFRNNATI-----DOWNAKNSDITVGQYD--QLVTR 124  
 DB 397 TVAAYDWIADFQKGDID----KIDLSAFRNEQQLSFQDQFTKGQEVML-QWDAANSITN 451  
 QY 125 VVTHEMAHASV--MYRQVG 141  
 DB 452 LMLHEAGHSSVDFLVRVIG 470  
 RESULT 12  
 ID GP63 LEIME STANDARD; PRT; 645 AA.  
 AC P43150;  
 DT 01-NOV-1995 (Rel. 32, Created)  
 DT 01-NOV-1995 (Rel. 32, Last sequence update)  
 DT 10-OCT-2003 (Rel. 42, Last annotation update)  
 DE Leishmanolysin Cl precursor (EC 3.4.24.36) (Cell surface protease)  
 DE (Major surface glycoprotein) (GP63 protein) (Promastigote surface  
 DE endopeptidase).  
 GN GP63-Cl.  
 OS Leishmania mexicana.  
 OC Eukaryota; Eulenzozoa; Kinetoplastida; Trypanosomatidae; Leishmania.  
 OX NCBI\_TaxID=5665;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=MNYC/BZ/62/M379;  
 RX MEDLINE=93149206; PubMed=8426614;  
 RA Medina-Acosta E.; Karsess R.E.; Russell D.G.;  
 RT "Structurally distinct genes for the surface protease of Leishmania  
 RT mexicana are developmentally regulated.";  
 RL Mol. Biochem. Parasitol. 57:31-46(1993).  
 CC -!- FUNCTION: Has an integral role during the infection of macrophages  
 CC in the mammalian host.  
 CC -!- CATALYTIC ACTIVITY: Preference for hydrophobic residues at P1 and  
 CC P1', and basic residues at P2 and P3'. A model nonapeptide is  
 CC cleaved at -Ala-Tyr-|-Leu-Lys-Lys-.  
 CC -!- COFACTOR: Binds 1 zinc ion per subunit (By similarity).  
 CC -!- DEVELOPMENTAL STAGE: Expressed in both the promastigote and the

```

CC anastigote forms.
CC -1- SIMILARITY: Belongs to peptidase family M8.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; X64394; CAA45733.1; --
DR PIR; S19916; S19916.
DR HSP; P08148; ILMU.
DR MEROPS; M08.001; --
DR GlycosuitedB; P43150; --
DR InterPro; IPR006025; Pept_M_Zn_BS.
DR InterPro; IPR001577; Peptidase_M8.
DR Pfam; PF01457; Peptidase_M8; 1.
DR PRINTS; PR00782; LSHMANOLYSIN.
DR PROSITE; PS00142; ZINC_PROTEASE; 1.
KW Hydrolase, Metalloprotease, Glycoprotein, Metal-binding, Zinc;
KW Zymogen, Signal, Cell adhesion, Multigene family.
FT SIGNAL 1 39
FT PROPEP 40 102 ACTIVATION PEPTIDE (POTENTIAL).
FT CHAIN 103 646 LESHMANOLYSIN C1.
FT METAL 266 265 ZINC (CATALYTIC) (BY SIMILARITY).
FT ACT_SITE 267 267 BY SIMILARITY.
FT METAL 270 270 ZINC (CATALYTIC) (BY SIMILARITY).
FT METAL 336 336 ZINC (CATALYTIC) (BY SIMILARITY).
FT DISULFID 127 144 BY SIMILARITY.
FT DISULFID 193 232 BY SIMILARITY.
FT DISULFID 316 388 BY SIMILARITY.
FT DISULFID 395 458 BY SIMILARITY.
FT DISULFID 408 427 BY SIMILARITY.
FT DISULFID 417 492 BY SIMILARITY.
FT DISULFID 469 513 BY SIMILARITY.
FT DISULFID 518 568 BY SIMILARITY.
FT DISULFID 538 561 BY SIMILARITY.
FT CARBOHYD 86 86 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 297 297 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 399 399 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 409 409 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 433 433 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 445 445 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 466 466 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 501 501 N-LINKED (GLCNAC. .) (POTENTIAL).
FT SEQUENCE 646 AA; 63054 MW; FE446DDC78C10B0A CRC64;
Query Match 11.2%; Score 87; DB 1; Length 646;
Best Local Similarity 73.1%; Pred. No. 3.9;
Matches 19; Conservative 1; Mismatches 0; Indels 6; Gaps 1;
QY 117 QYDQVTRVVTHEMAHASVWVROGVF 142
DB 255 RYDQVTRVVTHEMAHA-----VGF 274
RESULT 13
YF48 MYCTU
ID YF48 MYCTU STANDARD; PRT; 678 AA.
AC Q10778;
DT 01-OCT-1996 (Rel. 34, Created)
DT 15-JUL-1999 (Rel. 38, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Hypothetical PPE-family protein Rv1548c/MT1599.
GN Rv1548C OR MT1599 OR MTCY48.17.
OS Mycobacterium tuberculosis.
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC Corynebacterinaceae; Mycobacteriaceae; Mycobacterium.
CX NCBI_taxid=1773;
RN [1]
RP SEQUENCE FROM N.A.

```

```

RC STRAIN=H37RV;
RX MEDLINE=98295987; PubMed=9634230;
RA Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,
RA Gordon S.V., Eiglmeier K., Gas S., Barry C.E. III, Tekala F.,
RA Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,
RA Davies R., Devlin K., Feltwell T., Gentles S., Hamlin N., Holtroyd S.,
RA Hornsby T., Jagels K., Krogh A., McLean J., Moule S., Murphy L.,
RA Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,
RA Rutter S., Seeger K., Skelton S., Squares S., Squares R.,
RA Sultun J.E., Taylor K., Whitehead S., Barrell B.G.;
RT "Deciphering the biology of Mycobacterium tuberculosis from the
RT complete genome sequence.";
RL Nature 393:537-544 (1998).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=CDC 1551 / Oshkosh;
RX MEDLINE=22206494; PubMed=12218036;
RA Fleischmann R.D., Alland D., Eisen J.A., Carpenter L., White O.,
RA Peterson J., DeBoy R., Dodson R., Gwinn M., Haft D., Hickey E.,
RA Kolonay J.F., Nelson W.C., Umayam L.A., Ermolaeva M., Salzberg S.L.,
RA Delcher A., Utterback T., Weidman J., Khouri H., Gill J., Mikula A.,
RA Bishai W., Jacobs W.R. Jr., Venter J.C., Fraser C.M.;
RT "Whole-genome comparison of Mycobacterium tuberculosis clinical and
RT laboratory strains.";
RL J. Bacteriol. 184:5479-5490 (2002).
CC -1- SIMILARITY: Belongs to the mycobacterial PPE family.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; Z74020; CAA98335.1; --
DR EMBL; AE007026; AAK45866.1; ALT_INIT.
DR PIR; A70762; A70762.
DR TIGR; MT1599; --
DR TubercuList; Rv1548c; --
DR InterPro; IPR000030; Microbac_PPE.
DR InterPro; IPR002989; Mycobac_Pentapep.
DR Pfam; PF01469; Pentapeptide_2; 1.
DR Pfam; PF00823; PPE; 1.
KW Hypothetical protein; Transmembrane; Complete proteome.
FT TRANSMEM 14 34 POTENTIAL.
FT TRANSMEM 180 200 POTENTIAL.
FT CONFLICT 258 258 D -> G (IN REF. 2).
FT SEQUENCE 678 AA; 66736 MW; 209F1593D52533A2 CRC64;
Query Match 11.2%; Score 87; DB 1; Length 678;
Best Local Similarity 26.2%; Pred. No. 4.1;
Matches 34; Conservative 10; Mismatches 40; Indels 46; Gaps 7;
QY 26 WGGG--GNHN--GGN--GGSGPDSITLIVQYGSANAALALQSDARSETTITQSGYGNAGDVG 82
DB 267 WGSNIGSYNLGGN-----LGSYNLGSGN-----TGDNTFGGNTGNLN 306
QY 83 QGADNSTIELTQN---GFRNNATIDWNKNSDITVGYDQVTRVVTHEMAHASVWVRO 139
DB 307 VGGGN-----TGSNFGFGTGNVFGNGTGTDTNFGS-----GNLGSNG 346
QY 140 VGCGNATAN 149
DB 347 IGFNGKGSNN 356
RESULT 14
GP63 LEICH
ID GP63 LEICH STANDARD; PRT; 599 AA.
AC P15706;
DT 01-APR-1990 (Rel. 14, Created)
DT 01-APR-1990 (Rel. 14, Last sequence update)

```

10-OCT-2003 (Rel. 42, Last annotation update)  
 DE Leishmanolysin precursor (EC 3.4.24.36) (Cell surface protease)  
 DE (Major surface glycoprotein) (GP63 protein) (Promastigote surface  
 DE endopeptidase).  
 GN GP63.  
 OS Leishmania chagasi.  
 OC Eukaryota; Euglenozoa; Kinetoplastida; Trypanosomatidae; Leishmania.  
 OX NCBI\_TaxID=44271;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=90205976; PubMed=2320059;  
 RA Miller R.A., Reed S.G., Parsons M.;  
 RT "Leishmania gp63 molecule implicated in cellular adhesion lacks an  
 RT Arg-Gly-Asp sequence.";  
 RL Mol. Biochem. Parasitol. 39:267-274(1990).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=92112918; PubMed=1370484;  
 RA Ramancorothy R., Doneison J.E., Paetz K.E., Maybodi M., Roberts S.C.,  
 RA Wilson M.E.;  
 RT "Three distinct RNAs for the surface protease gp63 are differentially  
 RT expressed during development of Leishmania donovani chagasi  
 RT promastigotes to an infectious form.";  
 RL J. Biol. Chem. 267:1898-1895(1992).  
 CC -1- FUNCTION: Has an integral role during the infection of macrophages  
 CC in the mammalian host.  
 CC -1- CATALYTIC ACTIVITY: Preference for hydrophobic residues at P1 and  
 CC P1' and basic residues at P2 and P3'. A model nonapeptide is  
 CC cleaved at -Ala-Tyr|-Leu-Lys-Lys-  
 CC -1- COFACTOR: Binds 1 zinc ion per subunit (By similarity).  
 CC -1- SUBCELLULAR LOCATION: Attached to the membrane by a GPI-anchor.  
 CC -1- SIMILARITY: Belongs to peptidase family M8.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@sib-sib.ch](mailto:license@sib-sib.ch)).  
 CC -----  
 DR EMBL; M80672; AAA29238.1; -;  
 DR EMBL; M28527; AAA29235.1; -;  
 DR PIR; A44951; A44951.  
 DR HSP; P08148; ILM.  
 DR MEROPS; M08.001; -;  
 DR InterPro; IPR006025; Pept\_M\_Zn\_BS.  
 DR InterPro; IPR001577; Peptidase\_M8.  
 DR Pfam; PF01457; Peptidase\_M8; 1.  
 DR PRINTS; PR00782; LSHMANOLYSIN.  
 DR PROSITE; PS00142; ZINC\_PROTEASE; 1.  
 KW Hydrolase; Metalloprotease; Glycoprotein; Metal-binding; Zinc;  
 KW Zymogen; Signal; Cell adhesion; GPI-anchor; Lipoprotein.  
 FT SIGNAL 1 39  
 FT PROPEP 40 97  
 FT CHAIN 98 574  
 FT PROPEP 575 599  
 FT METAL 261 261  
 FT ACT\_SITE 262 262  
 FT METAL 265 265  
 FT METAL 331 331  
 FT DISULFID 122 139  
 FT DISULFID 188 227  
 FT DISULFID 311 383  
 FT DISULFID 390 452  
 FT DISULFID 403 422  
 FT DISULFID 412 486  
 FT DISULFID 463 507  
 FT DISULFID 512 562  
 FT DISULFID 532 555  
 FT CARBOHYD 297 297  
 FT CARBOHYD 394 394  
 FT LIPID 574 574

FT SEQUENCE 599 AA; 63848 MW; 746730AE8E2A2E7C CRC64;  
 SQ similarity).  
 Query Match 11.1%; Score 86; DB 1; Length 599;  
 Best Local Similarity 66.7%; Pred. No. 4.3;  
 Matches 20; Conservative 2; Mismatches 6; Indels 2; Gaps 1;  
 QY 117 QYDQLVTRVYVTHEMAHASVVMVRQVGFQGNNA 146  
 :|||||  
 DB 250 RYDQLVTRVYVTHEMAHA--LGFSVGFQFEGA 277  
 RESULT 15  
 AGOI ARATH  
 ID AGOI ARATH STANDARD; PRT; 1048 AA.  
 AC 004379; Q9LDP83;  
 DT 16-OCT-2001 (Rel. 40, Created)  
 DT 16-OCT-2001 (Rel. 40, Last sequence update)  
 DT 15-MAR-2004 (Rel. 43, Last annotation update)  
 DE Argonaute protein.  
 GN AGOI OR ATIG48410 OR F11A17.3 OR TLN15.2.  
 OS Arabidopsis thaliana (Mouse-ear cress).  
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids;  
 OC eurosids II; Brassicales; Brassicaceae; Arabidopsids.  
 OX NCBI\_TaxID=3702;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=sv. Columbia; TISSUE=Leaf;  
 RX MEDLINE=98090460; PubMed=9427751;  
 RA Bohmert K., Camus I., Bellini C., Bouchez D., Caboche M., Benning C.;  
 RT "AGO1 defines a novel locus of Arabidopsis controlling leaf  
 RT development.";  
 RL EMBL J. 17:170-180(1998).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=sv. Columbia;  
 RX MEDLINE=21016719; PubMed=11130712;  
 RA Theologis A., Ecker J.R., Palm C.J., Federspiel N.A., Kaul S., S.Y.,  
 RA White O., Alonso J., Altar J., Araujo R., Bowman C.L., Brooks S.Y.,  
 RA Buehler E., Chan A., Chao Q., Chen H., Cheuk R.F., Chin C.W.,  
 RA Chung M.K., Conn L., Conway A.B., Conway A.R., Creasy T.H., Dewar K.,  
 RA Dunn P., Ectup P., Feldblyum T.V., Feng J.-D., Fong B., Fujii C.Y.,  
 RA Gill J.E., Goldsmith A.D., Haas B., Hansen N.F., Hughes B., Huizar L.,  
 RA Hunter J.L., Jenkins J., Johnson-Hopson C., Khan S., Khaykin E.,  
 RA Kim C.J., Koo H.L., Kremenetskaia I., Kurtz D.B., Kwan A., Lam B.,  
 RA Langin-Hooper S., Lee A., Lee J.M., Lenz C.A., Li J.H., Li Y.-P.,  
 RA Lin X., Liu S.X., Liu Z.A., Luros J.S., Maiti R., Marziani A.,  
 RA Militscher J., Miranda M., Nguyen M., Nierman W.C., Osborne B.I.,  
 RA Pai G., Peterson J., Pham P.K., Rizzo M., Rooney T., Rowley D.,  
 RA Sakano H., Salzberg S.L., Schwartz J.R., Shinn P., Southwick A.M.,  
 RA Sun H., Tallon L.J., Tambunga G., Toriumi M.J., Town C.D.,  
 RA Utterback T., Van Aken S., Vaysberg M., Vysotskaia V.S., Walker M.,  
 RA Wu D., Yu G., Fraser C.M., Venter J.C., Davis R.W.;  
 RT "Sequence and analysis of chromosome 1 of the plant Arabidopsis  
 RT thaliana.";  
 RL Nature 408:816-820(2000).  
 RN [3]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=sv. Columbia;  
 RX MEDLINE=22954850; PubMed=14593172;  
 RA Yamada K., Lim J., Dale J.M., Chen H., Shinn P., Palm C.J.,  
 RA Southwick A.M., Wu H.C., Kim C.J., Nguyen M., Pham P.K., Cheuk R.F.,  
 RA Karlin-Newmann G., Liu S.X., Lam B., Sakano H., Wu T., Yu G.,  
 RA Miranda M., Quach H.L., Tripp M., Chang C.H., Lee J.M., Toriumi M.J.,  
 RA Chan M.M., Tang C.C., Onodera C.S., Deng J.M., Akiyama K., Ansari Y.,  
 RA Arakawa T., Barth J., Banno F., Bowser L., Brooks S.Y., Carninci P.,  
 RA Chao Q., Choy N., Eju A., Goldsmith A.D., Gurjani M., Hansen N.F.,  
 RA Hayashizaki Y., Johnson-Hopson C., Hsuan V.W., Iida K., Karnes M.,  
 RA Khan S., Koesena E., Ishida J., Jiang P.X., Jones T., Kawai J.,  
 RA Kamiya A., Meyers C., Nakajima M., Narusaka M., Seki M., Sakurai T.,  
 RA Satou M., Tamse R., Vaysberg M., Wallender E.K., Wong C., Yamamura Y.,  
 RA Yuan S., Shinozaki K., Davis R.W., Theologis A., Ecker J.R.;

```

RT "Empirical analysis of transcriptional activity in the Arabidopsis
RL Genome.";
RL Science 302:842-846(2003).
CC -!- FUNCTION: Essential for proper development of leaves and floral
CC organs, and formation of axillary meristems.
CC -!- SUBCELLULAR LOCATION: Cytoplasmic (Potential).
CC -!- ALTERNATIVE PRODUCTS:
CC Event=Alternative splicing; Named isoforms=1;
CC Comment=A number of isoforms are produced. According to EST
CC sequences;
CC Name=1;
CC IsoId=O04379-1; Sequence=Displayed;
CC -!- TISSUE SPECIFICITY: Widely expressed at low levels.
CC -!- DEVELOPMENTAL STAGE: Expressed throughout all developmental
CC stages.
CC -!- SIMILARITY: Belongs to the argonaute family.
CC -!- SIMILARITY: Contains 1 PAZ domain.
CC -!- SIMILARITY: Contains 1 Piwi domain.
CC -!- CAUTION: Ref.2 (AAF79718) sequence differs from that shown due to
CC erroneous gene model prediction.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; U91995; AAC18440.1; --
CC EMBL; AC007932; AAD49755.1; --
CC EMBL; AC020889; AAF79718.1; ALT_SEQ.
CC EMBL; BT000941; AAN41341.1; --
CC InterPro; IPR003100; PAZ.
CC InterPro; IPR003185; Piwi.
CC Pfam; PF02170; PAZ; 1.
CC Pfam; PF02171; Piwi; 1.
CC PROSITE; PS50821; PAZ; 1.
CC PROSITE; PS50822; PIWI; 1.
CC Developmental protein; Alternative splicing.
CC FT DOMAIN 391 501 PAZ.
CC FT DOMAIN 676 997 PIWI.
CC FT DOMAIN 13 104 GLY-RICH.
CC SEQUENCE 1048 AA; 116190 MW; 3B5146343A09C541 CRC64;
CC -----
Query Match 11.0%; Score 85.5; DB 1; Length 1048;
Best Local Similarity 25.7%; Pred.No.9.2;
Matches 36; Conservative 17; Mismatches 60; Indels 27; Gaps 3;

QY 27 GGGNGHNGGNSGSPDSTLTI-----YOGSANAALALQSDARKSETTITOSGYG 76
DB 95 GGYGGGGRGGSPSGPQSQVPELHQATSTPTQAVSSQPTLSEVSTQVPEPTVLAQQF- 153
QY 77 NGADVCOGADNSTIE-----LTQNGFRNNTIDQNAKNSDITVGCYDQ 120
DB 154 EQLSVEQGPSQAIQIPSSKAFKPMRPGKQSGKRCIVKANHFPELPDKDLHHYDV 213
QY 121 LVTRVVTTHMAHASVMVRQV 140
DB 214 TITPEVTSRVNPAVMKQLV 233

Search completed: August 2, 2004, 14:49:28
Job time : 6.3 secs

```



GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: August 2, 2004, 14:39:17 ; Search time 29.7 Seconds  
(without alignments)  
1604.150 Million cell updates/sec

Title: US-09-543-407-14  
Perfect score: 775  
Sequence: 1 MKLLKVAAPAAIVVSGSALA.....HASVNVQRQVGFNNATANYQ 151

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 1017041 seqs, 315518202 residues

Total number of hits satisfying chosen parameters: 1017041

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : SPTREMBL\_25:  
1: sp\_archaea:\*  
2: sp\_bacteria:\*  
3: sp\_fungi:\*  
4: sp\_human:\*  
5: sp\_invertebrate:\*  
6: sp\_mammal:\*  
7: sp\_mhc:\*  
8: sp\_organelle:\*  
9: sp\_phase:\*  
10: sp\_plant:\*  
11: sp\_rodent:\*  
12: sp\_virus:\*  
13: sp\_vertebrate:\*  
14: sp\_unclassified:\*  
15: sp\_virus:\*  
16: sp\_bacteriaph:\*  
17: sp\_archaea:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	684	88.3	152	2 O33802	O33802 salmonella
2	578.5	74.6	150	2 Q7X243	Q7X243 citrobacter
3	533	68.8	149	2 Q7X240	Q7X240 citrobacter
4	503.5	65.0	152	16 Q8CW63	Q8CW63 escherichia
5	438.5	56.6	150	2 Q7X237	Q7X237 enterobacte
6	385	49.7	76	2 Q54069	Q54069 salmonella
7	122	15.7	29	2 Q9S3U5	Q9S3U5 escherichia
8	109	14.1	179	2 Q33801	Q33801 salmonella
9	107.5	13.9	151	2 Q7X238	Q7X238 enterobacte
10	105	13.5	139	16 Q8EIH3	Q8EIH3 shewanella
11	104.5	13.5	152	2 Q7X241	Q7X241 citrobacter
12	104.5	13.5	160	16 Q8CW64	Q8CW64 escherichia
13	104	13.4	502	16 Q8EIH4	Q8EIH4 shewanella
14	103.5	13.4	151	2 Q7X244	Q7X244 citrobacter
15	103.5	13.4	151	16 Q7UCZ1	Q7UCZ1 shigella fl
16	103.5	13.4	160	16 Q83RU7	Q83RU7 shigella fl

17	101	13.0	362	16 Q8EV84	Q8EV84 mycoplasma
18	93.5	12.8	348	13 Q93397	Q93397 cyprinus ca
19	96.5	12.5	552	16 P96840	P96840 mycobacteri
20	96.5	12.5	552	16 Q7TW76	Q7TW76 mycobacteri
21	96.5	12.5	623	16 Q8VIY0	Q8VIY0 mycobacteri
22	96.5	12.5	1422	16 Q8EFU3	Q8EFU3 shewanella
23	96.5	12.5	3501	16 Q8Y106	Q8Y106 ralstonia s
24	96.5	12.5	3552	16 Q8XSD6	Q8XSD6 ralstonia s
25	96	12.4	154	16 Q8YJ15	Q8YJ15 bradyrhizob
26	96	12.4	157	16 Q88HG0	Q88HG0 pseudomonas
27	96	12.4	606	3 Q8P319	Q8P319 neurospora
28	95.5	12.3	624	3 Q8NIV1	Q8NIV1 neurospora
29	95.5	12.3	1765	16 Q7V8S5	Q7V8S5 prochloroco
30	95	12.3	1748	5 Q94821	Q94821 tetrathymena
31	94.5	12.2	191	3 Q8TF66	Q8TF66 saccharomyc
32	94	12.1	171	16 Q8YJ13	Q8YJ13 bradyrhizob
33	94	12.1	368	16 Q8EWD6	Q8EWD6 mycoplasma
34	93	12.0	480	16 Q89EV2	Q89EV2 bradyrhizob
35	93	12.0	2174	16 Q92U08	Q92U08 rhizobium m
36	92	11.9	453	5 Q9N6M8	Q9N6M8 drosophila
37	91	11.7	301	10 Q9XIL0	Q9XIL0 arabidopsis
38	91	11.7	453	5 Q9NGF6	Q9NGF6 drosophila
39	91	11.7	453	5 Q9NGF7	Q9NGF7 drosophila
40	91	11.7	1222	5 Q9W4F0	Q9W4F0 drosophila
41	91	11.7	1222	5 Q8T8L9	Q8T8L9 drosophila
42	91	11.7	1351	5 Q8SX56	Q8SX56 drosophila
43	91	11.7	1354	5 Q8MPN4	Q8MPN4 drosophila
44	90.5	11.7	145	16 Q8UGN9	Q8UGN9 agrobacteri
45	90.5	11.7	191	10 Q7XDR3	Q7XDR3 oryza sativ

ALIGNMENTS

RESULT 1

O33802 PRELIMINARY; PRT; 152 AA.  
ID O33802  
AC O33802; TREMBL:rel. 05, Created  
DT 01-JAN-1998 (TREMBL:rel. 05, Last sequence update)  
DT 01-JAN-1998 (TREMBL:rel. 05, Last sequence update)  
DT 01-DEC-2001 (TREMBL:rel. 19, Last annotation update)  
DE Agfa protein (Fragment).  
GN AGFA.  
OS Salmonella typhimurium.  
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;  
OC Enterobacteriaceae; Salmonella.  
OX NCBI\_TaxID=602;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=98053981; PubMed=9393832;  
RA Sukupolvi S.S., Lorentz R.G., Gordon J.I., Bian Z., Pfeifer J.D.,  
RA Normark S.U., Rhen M.;  
RT "Expression of thin, aggregative fimbriae promotes interaction of  
RT Salmonella typhimurium SR-11 with mouse small intestinal epithelial  
RT cells";  
RL Infect. Immun. 65:5320-5325(1997).  
DR EMBL; AJ000514; CAA04151.1;  
FT NON\_TER  
SQ SEQUENCE 152 AA; 15401 MW; 9DA7DADC2364E006 CRC64;

Query Match 88.3%; Score 684; DB 2; Length 152;

Best Local Similarity 89.4%; Pred. No. 1.2e-47;

Matches 135; Conservative 4; Mismatches 12; Indels 0; Gaps 0;

QY	1	MKLLKVAAPAAIVVSGSALA	GVVPGWGGGHHNGGSSGPDSTLITYQYGSANAALAQ	60
DB	1	MKLLKVAAPAAIVVSGSALA	GVVPGWGGGHHNGGSSGPDSTLITYQYGSANAALAQ	60
QY	61	SDARKSETTITGSGYGNAD	VGQGDADNSFIETQGFNNATIDQWNAKNSDITVGYDQ	120
DB	61	SDARKSETTITGSGYGNAD	VGQGDADNSFIETQGFNNATIDQWNAKNSDITVGYDQ	120
QY	121	LIVRVVTHEMAHASVNV	QRQVGFNNATANYQ	151



```

OX NCBI_TaxID=562;
RN (1)
RP SEQUENCE FROM N.A.
RC TRANSPOS-Insertion sequence Isl;
RX MEDLINE=99314153; PubMed=10386375;
RL La Ragione R.M., Collighan R.J., Woodward M.J.;
RT "Non-curliation of Escherichia coli O78:H80 isolates associated with
RL IS1 inserti on in cs8b and reduced persistence in poultry infection.";
RL FEMS Microbiol. Lett 175:247-253(1999).
DR EMBL; AJ131756; CAB45380.1; -.
RN NON_TER 29
SQ SEQUENCE 29 AA; 2789 MW; E29DFC07ABBB243 CRC64;

Query Match 15.7%; Score 122; DB 2; Length 29;
Best Local Similarity 89.7%; Pred. No. 0.003;
Matches 26; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 MKLIKVAAPFAIVVSGSALAGVVPQGGG 29
||| ||| ||| ||| ||| ||| ||| |||
DB 1 MKLIKVAIAIAIVFSGSALAGVVPQGG 29
||| ||| ||| ||| ||| ||| ||| |||

RESULT 8
O33801
ID O33801 PRELIMINARY; PRT; 179 AA.
AC O33801;
DT 01-JAN-1998 (TrEMBLrel. 05, Created)
DT 01-JAN-1998 (TrEMBLrel. 05, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE AgfB protein.
DE AGFB protein.
GN AGFB.
OS Salmonella typhimurium.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Salmonella.
OC NCBI_TaxID=602;
RN (1)
RP SEQUENCE FROM N.A.
RX MEDLINE=98053981; PubMed=939832;
RL Sukupolvi S.S., Lorentz R.G., Gordon J.I., Bian Z., Pfeifer J.D.,
RL Normark S.J., Rhen M.;
RT "Expression of thin, aggregative fimbriae promotes interaction of
RT Salmonella typhimurium SR-11 with mouse small intestinal epithelial
RT cells.";
RL Infect. Immun. 65:5320-5325(1997).
RL EMBL; AJ000514; CAA04150.1; -.
SQ SEQUENCE 179 AA; 19318 MW; A2BCCB648B3C0S0B CRC64;

Query Match 14.1%; Score 109; DB 2; Length 179;
Best Local Similarity 26.5%; Pred. No. 0.29;
Matches 36; Conservative 20; Mismatches 48; Indels 32; Gaps 4;

QY 16 GSALAGVVPQGGGHHNGHSGSPDSTLSIYQVGSANALQSDARKSETITQSGY 75
||| ||| ||| ||| ||| ||| ||| |||
DB 63 GSKLLSVISQ-ERGGNNRAKVQAGNYPAYTEQGNAN-----DASISQSA 109
||| ||| ||| ||| ||| ||| ||| |||

QY 76 GNGADVGGADNSTIELTNGFNFNATIDQWAKNSDITV-----GOYDQLV 123
||| ||| ||| ||| ||| ||| ||| |||
DB 110 GNSAAI-----SAAIQKSGSNKANITQYGTAVVQKQSHMAIQANITQYGT 162
||| ||| ||| ||| ||| ||| ||| |||

QY 124 RVVTHEMAHASVMVRQ 139
||| ||| ||| ||| ||| ||| ||| |||
DB 163 AVVYQKQSHMAIRVTQ 178
||| ||| ||| ||| ||| ||| ||| |||

RESULT 9
QX238
ID QX238 PRELIMINARY; PRT; 151 AA.
AC QX238;
DT 01-OCT-2003 (TrEMBLrel. 25, Created)
DT 01-OCT-2003 (TrEMBLrel. 25, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Nucleation component of curlin monomers.
DE CSGB.

```

OS Enterobacter sakazakii.  
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;  
 OC Enterobacteriaceae; Enterobacter.  
 OX NCBI\_TaxID=28141;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=Fec39;  
 RA Zogaj X., Bokranz W., Nimtz M., Romling U.;  
 RT "Production of Cellulose and Curli Fimbriae by Members of the Family  
 Enterobacteriaceae Isolated from the Human Gastrointestinal Tract.";  
 RL Infect. Immun. 72:4151-4158 (2003).  
 DR EMBL; AJ515702; CAD56677.1; -.  
 SQ SEQUENCE 151 AA; 15985 MW; F0B82BD2A27882B7 CRC64;

Query Match 13.9%; Score 107.5; DB 2; Length 151;  
 Best Local Similarity 30.8%; Pred. No. 0.31;  
 Matches 36; Conservative 14; Mismatches 42; Indels 25; Gaps 4;

QY 10 AAIIVVSGSALAGVVPQGGGNGGNGSSGPDSTLSIYQYGSANAALQSDARKSETT 69

DB 58 AQIRQEGSKLLSVSQ-----DGAGNRARVD-----QSGTYNTAWIDQS-GNGNDAG 103

QY 70 ITOSGYNGADVGQGDNSTIETQNGFRNATIDOWNAKNSDITVGYDQDLVTRV 126

DB 104 ITQDGYNSA-----KIIQSGNRANITQYGTQKTAVVQKQSQMAIRVI 149

## RESULT 10

Q8EIH3 PRELIMINARY; PRT; 139 AA.

AC Q8EIH3;  
 DT 01-MAR-2003 (TrEMBLrel. 23, Created)  
 DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)  
 DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)  
 DE Minor curlin subunit CsgB, putative.  
 GN S00866.  
 OS Shewanella oneidensis.  
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Alteromonadales;  
 OC Alteromonadaceae; Shewanella.  
 OX NCBI\_TaxID=70863;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=MR-1;  
 RX MDLINE=2297686; PubMed=12368813;  
 RA Heidelberg J.F., Paulsen I.T., Nelson K.E., Gaidos E.J., Nelson W.C.,  
 RA Read T.D., Eisen J.A., Seshadri R., Ward N., Meche B., Clayton R.A.,  
 RA Meyer R.T., Teapin A., Scott J., Beanan M., Brinkac L., Daugherty S.,  
 RA DeBoy R.T., Dodson R.J., Durkin A.S., Haft D.H., Kolonay J.F.,  
 RA Madupu R., Peterson J.D., Umayam L.A., White O., Wolf A.M.,  
 RA Vamathevan J., Weidman J., Imprim M., Lee K., Berry K., Lee C.,  
 RA Mueller J., Khouri H., Gill J., Utterback T.R., McDonald L.A.,  
 RA Feldblyum T.V., Smith H.O., Venter J.C., Neilson K.H., Fraser C.M.;  
 RT "Genome sequence of the dissimilatory metal ion-reducing bacterium  
 Shewanella oneidensis";  
 RL Nat. Biotechnol. 20:1118-1123 (2002).  
 DR EMBL; AE015532; AAN53942.1; -.  
 DR TIGR; S00866; -.  
 KW Complete proteome.  
 SQ SEQUENCE 139 AA; 14811 MW; 41EC1CFA76957920 CRC64;

Query Match 13.5%; Score 105; DB 16; Length 139;  
 Best Local Similarity 28.3%; Pred. No. 0.45;  
 Matches 32; Conservative 23; Mismatches 44; Indels 14; Gaps 4;

QY 39 SGPDSTLSIYQGSANAALQSDARKSETTITQSGYNGADVGQGDNSTIETQNGFR 98

DB 41 SGEDNLIDLVOQTANQGVIFQSGSDNS-AYYTQAGNDNISLVITQGTNNEVQLLOVQAQ 99

QY 99 NNATIDQWAKNSDITVGYDQDLVTRVTHEMAHASVWVRQGVFGNNATANY 151

DB 100 NKASITQ-----IGN-DNLVQ---LNLQSGNFGSTQIQADGAASITQY 139

## RESULT 11

Q7X241 PRELIMINARY; PRT; 152 AA.

AC Q7X241;  
 DT 01-OCT-2003 (TrEMBLrel. 25, Created)  
 DT 01-OCT-2003 (TrEMBLrel. 25, Last sequence update)  
 DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)  
 DE Nucleation component of curlin monomers.  
 GN CSGB.  
 OS Citrobacter freundii.  
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;  
 OC Enterobacteriaceae; Citrobacter.  
 OX NCBI\_TaxID=546;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=Fec4;  
 RA Zogaj X., Bokranz W., Nimtz M., Romling U.;  
 RT "Production of Cellulose and Curli Fimbriae by Members of the Family  
 Enterobacteriaceae Isolated from the Human Gastrointestinal Tract.";  
 RL Infect. Immun. 72:4151-4158 (2003).  
 DR EMBL; AJ515701; CAD56674.1; -.  
 SQ SEQUENCE 152 AA; 16149 MW; D063A527D45D4329 CRC64;

Query Match 13.5%; Score 104.5; DB 2; Length 152;  
 Best Local Similarity 29.1%; Pred. No. 0.55;  
 Matches 32; Conservative 13; Mismatches 40; Indels 25; Gaps 3;

QY 16 GSALAGVVPQGGGNGGNGSSGPDSTLSIYQYGSANAALQSDARKSETTITQSGY 75

DB 65 GSKLLSVISQ-----EGSGNRAKTDQT-----GSYNFAY-----IDQTGS 99

QY 76 GNGADVGQGDNSTIETQNGFRNATIDOWNAKNSDITVGYDQDLVTRV 125

DB 100 SNDASIKQSGYNGTAVIIQKSGNKANITQYGTQKTAVVQKQSQMAIRV 149

## RESULT 12

Q8CW64 PRELIMINARY; PRT; 160 AA.

AC Q8CW64;  
 DT 01-MAR-2003 (TrEMBLrel. 23, Created)  
 DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)  
 DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)  
 DE Minor curlin subunit precursor.  
 GN CSGB OR C1305.  
 OS Escherichia coli O6.  
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;  
 OC Enterobacteriaceae; Escherichia.  
 OX NCBI\_TaxID=217992;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=O6.H1 / CFT073 / ATCC 700928;  
 RX MEDLINE=2238234; PubMed=12471157;  
 RA Welch R.A., Burland V., Plunkett G. III, Redford P., Roesch P.,  
 RA Rasko D., Buckles E.L., Liou S.-R., Boutin A., Hackett J., Stroud D.,  
 RA Mayhew G.F., Rose D.J., Zhou S., Schwartz D.C., Perna N.T.,  
 RA Mobley H.L.T., Donnenberg M.S., Blattner F.R.;  
 RT "Extensive mosaic structure revealed by the complete genome sequence  
 of uropathogenic Escherichia coli";  
 RL Proc. Natl. Acad. Sci. U.S.A. 99:17020-17024 (2002).  
 DR EMBL; AE016759; AAN79778.1; -.  
 KW Complete proteome.  
 SQ SEQUENCE 160 AA; 16963 MW; 49F68448D979B986 CRC64;

Query Match 13.5%; Score 104.5; DB 16; Length 160;  
 Best Local Similarity 29.2%; Pred. No. 0.58;  
 Matches 35; Conservative 19; Mismatches 59; Indels 7; Gaps 4;

QY 9 FAAITVWGSAL--AGVVPQGGGNGGNGSSGPDSTLSIY-QYGSANAALQSDARK 65

DB 42 FAVNELSKSFFNQAAII--GQAGTNNNAQLRQGGSKLLTVAQGGSSNRA-KIDQTGY 97

QY 66 SETTITQSGYNGADVGQGDNSTIETQNGFRNATIDOWNAKNSDITVGYDQDLVTRV 125

Db 98 NLAYIDQAGSANDASISQAGYNTAMIIQKSGNKANITQYGTOKTAVVQSQMAIRV 157

RESULT 13  
Q8EIH4 PRELIMINARY; PRT; 502 AA.  
AC Q8EIH4;  
DT 01-MAR-2003 (T-EMBLrel. 23, Created)  
DT 01-MAR-2003 (T-EMBLrel. 23, Last sequence update)  
DT 01-MAR-2003 (T-EMBLrel. 23, Last annotation update)  
DE Conserved hypothetical protein.  
GN S00865.  
OS Shewanella oneidensis.  
OC Bacteria; Proteobacteria; Gammaproteobacteria; Alteromonadales;  
OC Alteromonadaceae; Shewanella.  
OX NCBI\_TaxID=70863;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=MR-1;  
RX MEDLINE=22297686; PubMed=12368813;  
RA Heidelberg J.F., Paulsen I.T., Nelson K.E., Caidos E.J., Nelson W.C.,  
RA Read T.D., Eisen J.A., Seshadri R., Ward N., Meche B., Clayton R.A.,  
RA Meyer T., Teapin A., Scott J., Beanan M., Brinkac L., Daugherty S.,  
RA DeBoy R.T., Dodson R.J., Durkin A.S., Haft D.H., Kolonay J.F.,  
RA Madupu R., Peterson J.D., Umayam L.A., White O., Wolf A.M.,  
RA Vamathevan J., Weidman J., Impraim M., Lee K., Berry K., Lee C.,  
RA Mueller J., Khouri H., Gill J., Utterback T.R., McDonald L.A.,  
RA Feldblyum T.V., Smith H.O., Venter J.C., Neallson K.H., Fraser C.M.;  
RT "Genome sequence of the dissimilatory metal ion-reducing bacterium  
Shewanella oneidensis";  
RL Nat. Biotechnol. 20:1118-1123 (2002).  
DR EMBL; AB015332; AAM53941.1; -.  
DR TIGR; S00865; -.  
KW Hypothetical protein; Complete proteome.  
SQ SEQUENCE 502 AA; 52441 MW; D08CA23D6C46B62D CRC64;

Query Match 13.4%; Score 104; DB 16; Length 502;  
Best Local Similarity 26.8%; Pred. No. 2.5; Indels 36; Gaps 7;  
Matches 40; Conservative 22; Mismatches 56;  
QY 29 GGNHNG-----GGN-----SSGPDSTLSIYQYGSANA---ALALOS 61  
DB 231 GDNHTGFVYALAGSENDISMEQSGNNAYLSTMTGDDNTVDITQDGSNTVGSLSIADI 290  
QY 62 DAKSETTITQSYGNGADVGQADNSTIELTQNGFRNNATIDOWNAKNS-DI-TVGYQYD 119  
DB 291 QGDDNDITIKQKGSNGAEFQVWGDSDNDVLDKRGDANFATFGAYGTDNDFDLSSKGDNN 350  
QY 120 QLVTRVVTTHMAHASVMVRQVGFEN-----NATAN 149  
DB 351 ELVAPATGED---NSTEISQEGDANFAYVDATGN 381

RESULT 14  
Q7X244 PRELIMINARY; PRT; 151 AA.  
AC Q7X244;  
DT 01-OCT-2003 (T-EMBLrel. 25, Created)  
DT 01-OCT-2003 (T-EMBLrel. 25, Last sequence update)  
DT 01-OCT-2003 (T-EMBLrel. 25, Last annotation update)  
DE Nucleation component of curlin monomers.  
GN CSGB.  
OS Citrobacter sp. Fec2.  
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;  
OC Enterobacteriaceae; Citrobacter.  
OX NCBI\_TaxID=213763;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=Fec2;  
RA Zogaj X., Bokranz W., Nimtz M., Romling U.;  
RT "Production of Cellulose and Curli Fimbriae by Members of the Family  
Enterobacteriaceae Isolated from the Human Gastrointestinal Tract.";

RL Infect. Immun. 72:4151-4158 (2003).  
DR EMBL; AJ515700; CAD56671.1; -.  
SQ SEQUENCE 151 AA; 16158 MW; BD00AF57E1400704 CRC64;

Query Match 13.4%; Score 103.5; DB 2; Length 151;  
Best Local Similarity 28.8%; Pred. No. 0.65; Indels 25; Gaps 3;  
Matches 32; Conservative 13; Mismatches 41;  
QY 16 GSALAGVVPQWGGGNGHNGSSGPDSTLSIYQYGSANAALALQSDARKSETTITQSGY 75  
DB 64 GSKLLSVVSQ-----EGGSNRAKVD-----QSGAYNFAY-----IAQSGH 98  
QY 76 GNGADVGQADNSTIELTQNGFRNNATIDOWNAKNSDITVGYDQYDQYDQYDQYD 126  
DB 99 SNDASISQSYNTAMIIQKSGNKANITQYGTOKTAVVQSQMAIRV 149

RESULT 15  
Q7UCZ1 PRELIMINARY; PRT; 151 AA.  
AC Q7UCZ1;  
DT 01-OCT-2003 (T-EMBLrel. 25, Created)  
DT 01-OCT-2003 (T-EMBLrel. 25, Last sequence update)  
DT 01-OCT-2003 (T-EMBLrel. 25, Last annotation update)  
DE Minor curlin subunit.  
GN CSGB OR S1108.  
OS Shigella flexneri.  
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;  
OC Enterobacteriaceae; Shigella.  
OX NCBI\_TaxID=623;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=2457T / ATCC 700930 / Serotype 2a;  
RX MEDLINE=22590274; PubMed=12704152;  
RA Wei J., Goldberg M.B., Burland V., Venkatesan M.M., Deng W.,  
RA Fournier G., Mayhew G.F., Plunkett G. III, Rose D.J., Darling A.,  
RA Mau B., Perna N.T., Payne S.M., Runyen-Janecky L.J., Zhou S.,  
RA Schwartz D.C., Blattner F.R.;  
RT "Complete genome sequence and comparative genomics of Shigella  
flexneri serotype 2a strain 2457T";  
RL Infect. Immun. 71:2775-2786 (2003).  
DR EMBL; AB016981; AAP16542.1; -.  
SQ SEQUENCE 151 AA; 15868 MW; 5D5D266B964014A0 CRC64;

Query Match 13.4%; Score 103.5; DB 16; Length 151;  
Best Local Similarity 28.3%; Pred. No. 0.65; Indels 7; Gaps 4;  
Matches 34; Conservative 20; Mismatches 59;  
QY 9 FAAIVVSGAL--AGVVPQWGGGNGHNGSSGPDSTLSIY-QYGSANAALALQSDARK 65  
DB 33 FAVNELSKSFPNQRAII---QAGTNNSAQLRQGGSKLLAVVQEGSSNRA-KIDQTGY 88  
QY 66 SETTITQSYGNGADVGQADNSTIELTQNGFRNNATIDOWNAKNSDITVGYDQYDQYD 125  
DB 89 NLAYIDQAGSANDASISQAGYNTAMIIQKSGNKANITQYGTOKTAVVQSQMAIRV 148

Search completed: August 2, 2004, 14:54:37  
Job time : 30.7 secs



GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: August 2, 2004, 14:35:42 ; Search time 44.9 Seconds

(without alignments)  
950.215 Million cell updates/sec

Title: US-09-543-407-16

Perfect score: 757

Sequence: 1 MKLLKVAFAAIVVSGSALA.....DSSWMVRQVFGNNATANY 151

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1586107 seqs, 282547505 residues

Total number of hits satisfying chosen parameters: 1586107

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : A Geneseqp29Jan04:\*

1: Geneseqp1980s:\*

2: Geneseqp1990s:\*

3: Geneseqp2000s:\*

4: Geneseqp2001s:\*

5: Geneseqp2002s:\*

6: Geneseqp2003as:\*

7: Geneseqp2003bs:\*

8: Geneseqp2004s:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	757	100.0	151	3 AAB36348	Aab36348 Agfa::PT3
2	658	86.9	151	2 AAR74625	Aar74625 Agfa sequ
3	658	86.9	151	3 AAB36341	Aab36341 Salmonell
4	653	86.3	151	2 AAW23570	Aaw23570 Salmonell
5	618	81.6	151	3 AAB36349	Aab36349 Agfa::PT3
6	580	76.6	151	3 AAB36346	Aab36346 Agfa::PT3
7	578	76.4	151	3 AAB36347	Aab36347 Agfa::PT3
8	577	76.2	151	3 AAB36353	Aab36353 Agfa::PT3
9	575	76.0	151	3 AAB36352	Aab36352 Agfa::PT3
10	574	75.8	151	3 AAB36350	Aab36350 Agfa::PT3
11	568	75.0	151	3 AAB36354	Aab36354 Agfa::PT3
12	567	74.9	151	3 AAB36351	Aab36351 Agfa::PT3
13	566	74.8	151	3 AAB36355	Aab36355 Agfa::PT3
14	487	64.3	120	2 AAR62761	Aar62761 Agfa sequ
15	487	64.3	120	2 AAW23569	Aaw23569 Salmonell
16	484	63.9	151	3 AAB36343	Aab36343 Escherich
17	479	63.3	151	7 ABR82651	Abr82651 E. coli C
18	406	53.6	142	2 AAR52664	Aar52664 Fibronect
19	345	45.6	122	2 AAR52663	Aar52663 FNB curli
20	237	31.3	45	3 AAB36316	Aab36316 Salmonell
21	123	16.2	23	3 AAB36321	Aab36321 Salmonell
22	123	16.2	23	3 AAB36326	Aab36326 Salmonell
23	123	16.2	23	3 AAB36338	Aab36338 Salmonell
24	118.5	15.7	151	3 AAB36344	Aab36344 Escherich
25	115	15.2	22	3 AAB36325	Aab36325 Salmonell

26	115	15.2	22	3 AAB36339	Aab36339 Salmonell
27	115	15.2	22	3 AAB36320	Aab36320 Salmonell
28	111.5	14.7	151	3 AAB36342	Aab36342 Salmonell
29	111	14.7	22	3 AAB36322	Aab36322 Salmonell
30	111	14.7	22	3 AAB36327	Aab36327 Salmonell
31	111	14.7	22	3 AAB36337	Aab36337 Salmonell
32	109	14.4	23	3 AAB36340	Aab36340 Salmonell
33	109	14.4	23	3 AAB36324	Aab36324 Salmonell
34	109	14.4	23	3 AAB36319	Aab36319 Salmonell
35	102	13.5	26	7 ABR82649	Abr82649 E. coli V
36	98	12.9	26	7 ABR82645	Abr82645 E. coli c
37	96.5	12.7	502	2 AAW32312	Aaw32312 Leishmani
38	96.5	12.7	2016	6 AAE36891	Aae36891 Plectreur
39	96	12.7	19	3 AAB36323	Aab36323 Salmonell
40	96	12.7	19	3 AAB36336	Aab36336 Salmonell
41	96	12.7	19	3 AAB36328	Aab36328 Salmonell
42	95	12.5	24	7 ABR82647	Abr82647 E. coli c
43	95	12.5	764	6 AAE36890	Aae36890 Plectreur
44	93	12.3	290	6 ABU44579	Abu44579 Protein e
45	92	12.2	23	3 AAB36331	Aab36331 Escherich

#### ALIGNMENTS

#### RESULT 1

AAB36348  
ID AAB36348 standard; protein; 151 AA.

XX

AC AAB36348;

XX

DT 26-FEB-2001 (first entry)

XX

DE Agfa::PT3#3 amino acid sequence SEQ ID NO:16.

XX

KW Salmonella; agfa; chromosomal gene replacement; fimbrin; epitope;

KW vaccine; immune response; immunogen.

XX

OS Salmonella enteritidis.

OS Escherichia coli.

OS Synthetic.

XX

PN WO200060102-A2.

XX

PD 12-OCT-2000.

XX

PF 05-APR-2000; 2000WO-CA000356.

XX

PR 05-APR-1999; 99US-0127888P.

XX

PA (UYVI-) UNIV VICTORIA.

XX

PI White AP, Doran JL, Collison SK, Kay WW;

XX

DR WPI; 2000-672631/65.

DR

N-PSDB; AAC64624.

XX

PS Recombinant agfa gene having a segment replaced by a foreign DNA sequence

XX

PT which encodes foreign epitope or antigen, expresses recombinant Agfa

XX

PT protein useful for eliciting immune response in animal.

XX

PS Disclosure; Page 136; 139pp; English.

XX

CC The present invention describes a recombinant agfa gene (I) where a

CC segment of the gene has been replaced by a segment of a foreign DNA

CC sequence which encodes a foreign epitope or antigen. Also described are:

CC (1) use of thin aggregative fimbriae (SEF17/TAf) nucleation depended

CC assembly system of strains of Salmonella, Escherichia coli and

CC Enterobacteriaceae for the production of fimbriae comprising recombinant

CC Agfa, CSga and Agfa-homologue fimbrian subunits, respectively; (2)

CC directing recombination of a recombinant gene into the chromosome of the

CC homologous species; (3) directing recombination of a recombinant gene

CC back into the chromosome of the homologous species, replacing the native

CC copy of that gene; and (4) eliciting an immune response in an animal,  
 CC comprising separating an amino acid polymer comprising a recombinant Agfa  
 CC protein containing a replacement segment or segments of foreign amino  
 CC acid sequence or sequences grown on a Salmonella, E. coli or  
 CC Enterobacteriaceae host cell, from the host cell and introducing the  
 CC polymer into the animal in conjunction with a carrier or diluent. (1) is  
 CC useful for the expression of recombinant Agfa protein which is useful for  
 CC eliciting an immune response in an animal. In a fimbrial presentation  
 CC system the heterologous antigens are presented in high numbers (up to  
 CC 500,000 copies/cell), the hybrid fimbriae protein possesses both the  
 CC immunogenicity and adhesion properties relevant for an efficient live  
 CC vaccine, the carrier fimbrial subunit proteins are usually strong  
 CC immunogens, which may be important for directing an immune response  
 CC against the inserted epitope, and hybrid fimbriae are easy and  
 CC inexpensive to purify in large amount. The present sequence is given in  
 CC the exemplification of the present invention

XX Sequence 151 AA;

Query Match 100.0%; Score 757; DB 3; Length 151;  
 Best Local Similarity 100.0%; Pred. No. 2.2e-67;  
 Matches 151; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MKLLKVAFAAIVVSGSALAGVVDQLVTRVVTHEMAHASGPDSTLSIYQYGSANAALALQ 60  
 DB 1 MKLLKVAFAAIVVSGSALAGVVDQLVTRVVTHEMAHASGPDSTLSIYQYGSANAALALQ 60

QY 61 SDARKSETTITQSGYNGADVGGADNSTIELTQNGFRNNATIDQWNAKNSDITVGYGG 120  
 DB 61 SDARKSETTITQSGYNGADVGGADNSTIELTQNGFRNNATIDQWNAKNSDITVGYGG 120

QY 121 NNAALVNOTASDSSVMVRQVGFNNATANQY 151  
 DB 121 NNAALVNOTASDSSVMVRQVGFNNATANQY 151

RESULT 2  
 AAR74625  
 ID AAR74625 standard; protein; 151 AA.  
 XX AC AAR74625;  
 XX DT 25-MAR-2003 (revised)  
 XX DT 26-JUN-1995 (first entry)  
 XX DE Agfa sequence.  
 XX KW Salmonella; Agfa; vaccine.  
 XX OS Salmonella.  
 XX PN WO9425598-A2.  
 XX FD 10-NOV-1994.  
 XX PF 26-APR-1994; 94WO-IB000207.  
 XX PR 26-APR-1993; 93US-00054452.  
 XX PA (UYVI-) UNIV VICTORIA INNOVATION & DEV CORP.  
 XX PA (KING/) KING J.  
 XX PI Kay WW, Collinson SK, Clouthier SC, Doran JL;  
 XX WPI; 1994-358275/44.  
 XX DR N-PSDB; AAQ87467.  
 XX CC Eliciting an immune response to Salmonella - using attenuated Salmonella  
 XX PT strains, vector constructs, or compens. contg. fimbrial type proteins.  
 XX PS Disclosure; Fig 7B; 95pp; English.  
 XX CC The Salmonella Agfa protein and DNA are used in vaccine and genetic

CC immunization compositions, respectively, to elicit an immune response to  
 CC Salmonella in animals (e.g. food producing animals) and humans. (Updated  
 CC on 25-MAR-2003 to correct PN field.)

XX Sequence 151 AA;

Query Match 86.9%; Score 658; DB 2; Length 151;  
 Best Local Similarity 90.1%; Pred. No. 1.6e-57;  
 Matches 136; Conservative 3; Mismatches 12; Indels 0; Gaps 0;

QY 1 MKLLKVAFAAIVVSGSALAGVVDQLVTRVVTHEMAHASGPDSTLSIYQYGSANAALALQ 60  
 DB 1 MKLLKVAFAAIVVSGSALAGVVDQLVTRVVTHEMAHASGPDSTLSIYQYGSANAALALQ 60

QY 61 SDARKSETTITQSGYNGADVGGADNSTIELTQNGFRNNATIDQWNAKNSDITVGYGG 120  
 DB 61 SDARKSETTITQSGYNGADVGGADNSTIELTQNGFRNNATIDQWNAKNSDITVGYGG 120

QY 121 NNAALVNOTASDSSVMVRQVGFNNATANQY 151  
 DB 121 NNAALVNOTASDSSVMVRQVGFNNATANQY 151

RESULT 3  
 AAB36341  
 ID AAB36341 standard; protein; 151 AA.

XX AC AAB36341;  
 XX DT 26-FEB-2001 (first entry)  
 XX DE Salmonella enteritidis Agfa amino acid sequence SEQ ID NO:5.  
 XX KW Salmonella; agfa; chromosomal gene replacement; fimbria; epitope;  
 XX KW vaccine; immune response; immunogen.  
 XX OS Salmonella enteritidis.

XX WO200060102-A2.

XX PD 12-OCT-2000.

XX PF 05-APR-2000; 2000WO-CA000356.

XX PR 05-APR-1999; 99US-0127888P.

XX PA (UYVI-) UNIV VICTORIA.

XX PI White AP, Doran JL, Collinson SK, Kay WW;

XX WPI; 2000-672631/65.

XX DR N-PSDB; AAC64617.

XX CC Recombinant agfa gene having a segment replaced by a foreign DNA sequence  
 XX PT which encodes foreign epitope or antigen, expresses recombinant Agfa  
 XX protein useful for eliciting immune response in animal.

XX Disclosure; Page 135; 139pp; English.

XX The present invention describes a recombinant agfa gene (I) where a  
 XX segment of the gene has been replaced by a segment of a foreign DNA  
 XX sequence which encodes a foreign epitope or antigen. Also described are:  
 XX (1) use of thin aggregative fimbriae (SEF17/TAF) nucleation depended  
 XX assembly system of strains of Salmonella, Escherichia coli and  
 XX Enterobacteriaceae for the production of fimbriae comprising recombinant  
 XX Agfa, CsgA and Agfa-homologue fimbria subunits, respectively; (2)  
 XX directing recombination of a recombinant gene into the chromosome of the  
 XX homologous species; (3) directing recombination of a recombinant gene  
 XX back into the chromosome of the homologous species, replacing the native  
 XX copy of that gene; and (4) eliciting an immune response in an animal,  
 XX comprising separating an amino acid polymer comprising a recombinant Agfa  
 XX protein containing a replacement segment or segments of foreign amino  
 XX acid sequence or sequences grown on a Salmonella, E. coli or



CC Enterobacteriaceae host cell, from the host cell and introducing the  
CC polymer into the animal in conjunction with a carrier or diluent. (1) is  
CC useful for the expression of recombinant Agfa protein which is useful for  
CC eliciting an immune response in an animal. In a fibrial presentation  
CC system the heterologous antigens are presented in high numbers (up to  
CC 500,000 copies/cell) the hybrid fimbria protein possesses both the  
CC immunogenicity and adhesion properties relevant for an efficient live  
CC vaccine, the carrier fibrial subunit proteins are usually strong  
CC immunogens, which may be important for directing an immune response  
CC against the inserted epitope, and hybrid fimbriae are easy and  
CC inexpensive to purify in large amount. The present sequence is given in  
CC the exemplification of the present invention

XX Sequence 151 AA;  
SQ

Query Match 86.9%; Score 658; DB 3; Length 151;  
Best Local Similarity 90.1%; Pred. No. 1.6e-57;  
Matches 136; Conservative 3; Mismatches 12; Indels 0; Gaps 0;

QY 1 MKLLKVAFAAIVVSGSALAGVYDQLVTRVVTHEMAHAGSPDSTLSIYQYGSANAALQ 60  
Db 1 MKLLKVAFAAIVVSGSALAGVYDQLVTRVVTHEMAHAGSPDSTLSIYQYGSANAALQ 60

QY 61 SDARKSETTITQSGYNGADVGQADNSTIELTQNGFRNNATIDQWNAKNSDITVGYGG 120  
Db 61 SDARKSETTITQSGYNGADVGQADNSTIELTQNGFRNNATIDQWNAKNSDITVGYGG 120

QY 121 NNAALVNOTASDSSVMVVRQVFGNNATANQY 151  
Db 121 NNAALVNOTASDSSVMVVRQVFGNNATANQY 151

RESULT 4  
AAW23570  
ID AAW23570 standard; protein; 151 AA.  
XX  
AC AAW23570;  
XX  
DT 25-MAR-2003 (revised)  
DT 29-SEP-1997 (first entry)  
DE Salmonella enteritidis 27655-3b agfa.  
KW Enteropathogenic bacteria; enterobacteria; S. enteritidis; antibody.  
XX Salmonella enteritidis.  
XX  
FH Key Location/Qualifiers  
FT Misc-difference 123  
FT /note= "Encoded by GCC"  
XX  
PN US5635617-A.  
XX  
PD 03-JUN-1997.  
XX  
PF 26-APR-1994; 94US-00233788.  
XX  
PR 26-APR-1993; 93US-00054452.  
XX  
PA (UUVI-) UNIV VICTORIA INNOVATION & DEV CORP.  
XX  
PI Collinson SK, Kay WW, Doran JL;  
XX WPI; 1997-309886/28.  
DR N-PSDB; AAT74142.  
XX  
PT Isolated Salmonella gene agfa - used for diagnosis of Salmonella or  
PT enteropathogenic bacteria of the Enterobacteria family.  
XX  
PS Example 2; Fig 7; 85pp; English.  
XX  
CC The present sequence represents agfa encoded by the full agfa gene  
CC derived from Salmonella enteritidis 27655-3b. The nucleic acid can be

CC used to provide diagnostic assays for Salmonella and/or enteropathogenic  
CC bacteria of the family Enterobacteria. It can also be used to provide  
CC proteins and antibodies which can be used for assays. The nucleic acid  
CC sequence can be used to provide probes or primers which can specifically  
CC hybridize to nucleic acid molecules from greater than 9% of Salmonella  
CC strains that are pathogenic to warm-blooded animals relative to nucleic  
CC acid molecules from virtually all other microbial organisms. (Updated on  
CC 25-MAR-2003 to correct PF field.)

XX Sequence 151 AA;  
SQ

Query Match 86.3%; Score 653; DB 2; Length 151;  
Best Local Similarity 89.4%; Pred. No. 5.1e-57;  
Matches 135; Conservative 3; Mismatches 13; Indels 0; Gaps 0;

QY 1 MKLLKVAFAAIVVSGSALAGVYDQLVTRVVTHEMAHAGSPDSTLSIYQYGSANAALQ 60  
Db 1 MKLLKVAFAAIVVSGSALAGVYDQLVTRVVTHEMAHAGSPDSTLSIYQYGSANAALQ 60

QY 61 SDARKSETTITQSGYNGADVGQADNSTIELTQNGFRNNATIDQWNAKNSDITVGYGG 120  
Db 61 SDARKSETTITQSGYNGADVGQADNSTIELTQNGFRNNATIDQWNAKNSDITVGYGG 120

QY 121 NNAALVNOTASDSSVMVVRQVFGNNATANQY 151  
Db 121 NNAALVNOTASDSSVMVVRQVFGNNATANQY 151

RESULT 5  
AAB36349  
ID AAB36349 standard; protein; 151 AA.  
XX  
AC AAB36349;  
XX  
DT 26-FEB-2001 (first entry)  
DE Agfa: PT3#4 amino acid sequence SEQ ID NO:18.  
XX  
KW Salmonella; agfa; chromosomal gene replacement; fimbria; epitope;  
KW vaccine; immune response; immunogen.  
XX  
OS Salmonella enteritidis.  
OS Escherichia coli.  
OS Synthetic.  
XX  
PN WO200060102-A2.  
XX  
PD 12-OCT-2000.  
XX  
PF 05-APR-2000; 2000WO-CA000356.  
XX  
PR 05-APR-1999; 99US-0127888P.  
XX  
PA (UUVI-) UNIV VICTORIA.  
XX  
PI White AP, Doran JL, Collinson SK, Kay WW;  
XX WPI; 2000-672631/65.  
DR N-PSDB; AAC64625.  
XX  
PT Recombinant agfa gene having a segment replaced by a foreign DNA sequence  
PT which encodes foreign epitope or antigen, expresses recombinant Agfa  
PT protein useful for eliciting immune response in animal.  
XX  
PS Disclosure; Page 136; 139pp; English.  
XX  
CC The present invention describes a recombinant agfa gene (1) where a  
CC segment of the gene has been replaced by a segment of a foreign DNA  
CC sequence which encodes a foreign epitope or antigen. Also described are:  
CC (1) use of thin aggregative fimbriae (SEF17/TAf) nucleation depended  
CC assembly system of strains of Salmonella, Escherichia coli and  
CC Enterobacteriaceae for the production of fimbriae comprising recombinant  
CC Agfa, CsgA and Agfa-homologue fimbria subunits, respectively; (2)

CC directing recombination of a recombinant gene into the chromosome of the  
 CC homologous species; (3) directing recombination of a recombinant gene  
 CC back into the chromosome of the homologous species, replacing the native  
 CC copy of that gene; and (4) eliciting an immune response in an animal,  
 CC comprising separating an amino acid polymer comprising a recombinant Agfa  
 CC protein containing a replacement segment or segments of foreign amino  
 CC acid sequence or sequences grown on a Salmonella, E. coli or  
 CC Enterobacteriaceae host cell, from the host cell and introducing the  
 CC polymer into the animal in conjunction with a carrier or diluent. (1) is  
 CC useful for the expression of recombinant Agfa protein which is useful for  
 CC eliciting an immune response in an animal. In a fimbrial presentation  
 CC system the heterologous antigens are presented in high numbers (up to  
 CC 500,000 copies/cell), the hybrid fimbrial protein possesses both the  
 CC immunogenicity and adhesion properties relevant for an efficient live  
 CC vaccine, the carrier fimbrial subunit proteins are usually strong  
 CC immunogens, which may be important for directing an immune response  
 CC against the inserted epitope, and hybrid fimbriae are easy and  
 CC inexpensive to purify in large amount. The present sequence is given in  
 CC the exemplification of the present invention  
 XX  
 XX Sequence 151 AA;

Query Match 81.6%; Score 618; DB 3; Length 151;  
 Best Local Similarity 76.6%; Pred. No. 1.6e-53;  
 Matches 131; Conservative 0; Mismatches 0; Indels 40; Gaps 2;  
 QY 1 MKLLKVAFAAIVVSGSALAGV-----YDQVTRVVTHEMAHAGS 40  
 Db 1 MKLLKVAFAAIVVSGSALAGVVPQGGGNGHNGSGDPDQVTRVVTHEMAHAGS 57  
 QY 41 PDSTLSIYQYGSANAALALQSDARKSETTITQSGYNGADVGQADNSTIQTQNGFRNN 100  
 Db 58 -----ALQSDARKSETTITQSGYNGADVGQADNSTIQTQNGFRNN 100  
 QY 101 ATTDQWNAKNSDITVQYGGNNAALVNOTASDSSVMVVRQVFGNNATANOY 151  
 Db 101 ATTDQWNAKNSDITVQYGGNNAALVNOTASDSSVMVVRQVFGNNATANOY 151

RESULT 6  
 AAB36346  
 ID AAB36346 standard; protein; 151 AA.  
 AC AAB36346;  
 XX  
 XX 26-FEB-2001 (first entry)  
 XX  
 XX Agfa::PT3#1 amino acid sequence SEQ ID NO:12.  
 XX  
 XX Salmonella; agfa; chromosomal gene replacement; fimbria; epitope;  
 KW vaccine; immune response; immunogen.  
 KW  
 XX Salmonella enteritidis.  
 OS Escherichia coli.  
 OS Synthetic.  
 XX  
 XX WO2000060102-A2.  
 XX  
 XX 12-OCT-2000.  
 PD  
 XX 05-APR-2000; 2000WO-CA000356.  
 PF  
 XX 05-APR-1999; 99US-0127888P.  
 XX  
 XX (UYVI-) UNIV VICTORIA.  
 PA  
 XX White AP, Doran JL, Collison SK, Kay WW;  
 PI  
 XX WPI; 2000-672631/65.  
 DR N-PSDB; AAC64622.  
 XX  
 XX Recombinant agfa gene having a segment replaced by a foreign DNA sequence  
 PT which encodes foreign epitope or antigen, expresses recombinant Agfa

PT protein useful for eliciting immune response in animal.  
 XX  
 XX Disclosure; Page 135; 139pp; English.  
 XX  
 XX The present invention describes a recombinant agfa gene (I) where a  
 CC segment of the gene has been replaced by a segment of a foreign DNA  
 CC sequence which encodes a foreign epitope or antigen. Also described are:  
 CC (1) use of thin aggregative fimbriae (SEF17/TAP) nucleation depended  
 CC assembly system of strains of Salmonella, Escherichia coli and  
 CC Enterobacteriaceae for the production of fimbriae comprising recombinant  
 CC Agfa, CsgA and Agfa-homologue fimbrial subunits, respectively; (2)  
 CC directing recombination of a recombinant gene into the chromosome of the  
 CC homologous species; (3) directing recombination of a recombinant gene  
 CC back into the chromosome of the homologous species, replacing the native  
 CC copy of that gene; and (4) eliciting an immune response in an animal,  
 CC comprising separating an amino acid polymer comprising a recombinant Agfa  
 CC protein containing a replacement segment or segments of foreign amino  
 CC acid sequence or sequences grown on a Salmonella, E. coli or  
 CC Enterobacteriaceae host cell, from the host cell and introducing the  
 CC polymer into the animal in conjunction with a carrier or diluent. (1) is  
 CC useful for the expression of recombinant Agfa protein which is useful for  
 CC eliciting an immune response in an animal. In a fimbrial presentation  
 CC system the heterologous antigens are presented in high numbers (up to  
 CC 500,000 copies/cell), the hybrid fimbrial protein possesses both the  
 CC immunogenicity and adhesion properties relevant for an efficient live  
 CC vaccine, the carrier fimbrial subunit proteins are usually strong  
 CC immunogens, which may be important for directing an immune response  
 CC against the inserted epitope, and hybrid fimbriae are easy and  
 CC inexpensive to purify in large amount. The present sequence is given in  
 CC the exemplification of the present invention  
 XX  
 XX Sequence 151 AA;

Query Match 76.6%; Score 580; DB 3; Length 151;  
 Best Local Similarity 80.1%; Pred. No. 9.8e-50;  
 Matches 121; Conservative 7; Mismatches 23; Indels 0; Gaps 0;  
 QY 1 MKLLKVAFAAIVVSGSALAGVVDQVTRVVTHEMAHAGSDSTLSIYQYGSANAALALQ 60  
 Db 1 MKLLKVAFAAIVVSGSALAGVVPQGGGNGHNGSGDPDSTLSIYQYGSANAALALQ 60  
 QY 61 SDARKSETTITQSGYNGADVGQADNSTIQTQNGFRNNATTDQWNAKNSDITVQYGG 120  
 Db 61 SDARKSETTITQSGYNGADVGQADNSTIQTQNGFRNNATTDQWNAKNSDITVQYGG 120  
 QY 121 NNAALVNOTASDSSVMVVRQVFGNNATANOY 151  
 Db 121 NNAALVNOTASDSSVMVVRQVFGNNATANOY 151

RESULT 7  
 AAB36347  
 ID AAB36347 standard; protein; 151 AA.  
 AC AAB36347;  
 XX  
 XX 26-FEB-2001 (first entry)  
 XX  
 XX Agfa::PT3#2 amino acid sequence SEQ ID NO:14.  
 DE  
 XX Salmonella; agfa; chromosomal gene replacement; fimbria; epitope;  
 KW vaccine; immune response; immunogen.  
 KW  
 XX Salmonella enteritidis.  
 OS Escherichia coli.  
 OS Synthetic.  
 XX  
 XX WO2000060102-A2.  
 XX  
 XX 12-OCT-2000.  
 PD  
 XX 05-APR-2000; 2000WO-CA000356.  
 PF  
 XX

```
PR 05-APR-1999; 99US-0127888P.
XX (UYVI-) UNIV VICTORIA.
XX White AP, Doran JL, Collison SK, Kay WW;
PI WPI; 2000-672631/65.
DR N-PSDB; AAC64623.
XX Recombinant agfa gene having a segment replaced by a foreign DNA sequence
PT which encodes foreign epitope or antigen, expresses recombinant Agfa
PT protein useful for eliciting immune response in animal.
XX Disclosure; Page 136; 139pp; English.
XX The present invention describes a recombinant agfa gene (I) where a
XX segment of the gene has been replaced by a segment of a foreign DNA
XX sequence which encodes a foreign epitope or antigen. Also described are:
CC (1) use of thin aggregative fimbriae (SEF17/TAF) nucleation depended
CC assembly system of strains of Salmonella, Escherichia coli and
CC Enterobacteriaceae for the production of fimbriae comprising recombinant
CC Agfa, CsgA and Agfa-homologue fimbrial subunits, respectively; (2)
CC directing recombination of a recombinant gene into the chromosome of the
CC homologous species; (3) directing recombination of a recombinant gene
CC back into the chromosome of the homologous species, replacing the native
CC copy of that gene; and (4) eliciting an immune response in an animal,
CC comprising separating an amino acid polymer comprising a recombinant Agfa
CC protein containing a replacement segment or segments of foreign amino
CC acid sequence or sequences grown on a Salmonella, E. coli or
CC Enterobacteriaceae host cell, from the host cell and introducing the
CC polymer into the animal in conjunction with a carrier or diluent. (I) is
CC useful for the expression of recombinant Agfa protein which is useful for
CC eliciting an immune response in an animal. In a fimbrial presentation
CC system the heterologous antigens are presented in high numbers (up to
CC 500,000 copies/cell), the hybrid fimbrial protein possesses both the
CC immunogenicity and adhesion properties relevant for an efficient live
CC vaccine, the carrier fimbrial subunit proteins are usually strong
CC immunogens, which may be important for directing an immune response
CC against the inserted epitope, and hybrid fimbriae are easy and
CC inexpensive to purify in large amount. The present sequence is given in
CC the exemplification of the present invention
XX Sequence 151 AA;
SQ Query Match 76.4%; Score 578; DB 3; Length 151;
Best Local Similarity 80.8%; Pred. No. 1.5e-49;
Matches 122; Conservative 6; Mismatches 23; Indels 0; Gaps 0;
QY 1 MKLLKVAFAAIVVSGSALAGVYDQVLTTRVVTTHMAHAGSPDSTLSIYQYGSANAALALQ 60
DB 1 MKLLKVAFAAIVVSGSALAGVYDQVLTTRVVTTHMAHAGSPDSTLSIYQYGSANAALALQ 60
QY 61 SDARKSETTITQSGYNGADVGGADNSTIELTQNGFRNNATIDQWNAKNSDITVGOYGG 120
DB 61 SDARKSETTITQSGYNGADVGGADNSTIELTQNGFRNNATIDQWNAKNSDITVGOYDQ 120
QY 121 NNAALVNQATSDSSVMVRQVGFNNATANQY 151
DB 121 LVTRVVTTHMAHAGSVVMVRQVGFNNATANQY 151
RESULT 8
AAB36353
ID AAB36353 standard; protein; 151 AA.
XX XX
AC AAB36353;
XX 26-FEB-2001 (first entry)
XX XX
DE Agfa::FT3#8 amino acid sequence SEQ ID NO:26.
XX Salmonella; agfa; chromosomal gene replacement; fimbrial; epitope;
KW vaccine; immune response; immunogen.
```

```
XX Salmonella enteritidis.
OS Escherichia coli.
OS Synthetic.
XX WO2000060102-A2.
XX 12-OCT-2000.
XX 05-APR-2000; 2000WO-CA000356.
XX 05-APR-1999; 99US-0127888P.
XX (UYVI-) UNIV VICTORIA.
XX White AP, Doran JL, Collison SK, Kay WW;
XX WPI; 2000-672631/65.
XX N-PSDB; AAC64629.
XX Recombinant agfa gene having a segment replaced by a foreign DNA sequence
XX which encodes foreign epitope or antigen, expresses recombinant Agfa
XX protein useful for eliciting immune response in animal.
XX Disclosure; Page 138; 139pp; English.
XX The present invention describes a recombinant agfa gene (I) where a
XX segment of the gene has been replaced by a segment of a foreign DNA
XX sequence which encodes a foreign epitope or antigen. Also described are:
CC (1) use of thin aggregative fimbriae (SEF17/TAF) nucleation depended
CC assembly system of strains of Salmonella, Escherichia coli and
CC Enterobacteriaceae for the production of fimbriae comprising recombinant
CC Agfa, CsgA and Agfa-homologue fimbrial subunits, respectively; (2)
CC directing recombination of a recombinant gene into the chromosome of the
CC homologous species; (3) directing recombination of a recombinant gene
CC back into the chromosome of the homologous species, replacing the native
CC copy of that gene; and (4) eliciting an immune response in an animal,
CC comprising separating an amino acid polymer comprising a recombinant Agfa
CC protein containing a replacement segment or segments of foreign amino
CC acid sequence or sequences grown on a Salmonella, E. coli or
CC Enterobacteriaceae host cell, from the host cell and introducing the
CC polymer into the animal in conjunction with a carrier or diluent. (I) is
CC useful for the expression of recombinant Agfa protein which is useful for
CC eliciting an immune response in an animal. In a fimbrial presentation
CC system the heterologous antigens are presented in high numbers (up to
CC 500,000 copies/cell), the hybrid fimbrial protein possesses both the
CC immunogenicity and adhesion properties relevant for an efficient live
CC vaccine, the carrier fimbrial subunit proteins are usually strong
CC immunogens, which may be important for directing an immune response
CC against the inserted epitope, and hybrid fimbriae are easy and
CC inexpensive to purify in large amount. The present sequence is given in
CC the exemplification of the present invention
XX Sequence 151 AA;
SQ Query Match 76.2%; Score 577; DB 3; Length 151;
Best Local Similarity 80.8%; Pred. No. 1.9e-49;
Matches 122; Conservative 5; Mismatches 24; Indels 0; Gaps 0;
QY 1 MKLLKVAFAAIVVSGSALAGVYDQVLTTRVVTTHMAHAGSPDSTLSIYQYGSANAALALQ 60
DB 1 MKLLKVAFAAIVVSGSALAGVYDQVLTTRVVTTHMAHAGSPDSTLSIYQYGSANAALALQ 60
QY 61 SDARKSETTITQSGYNGADVGGADNSTIELTQNGFRNNATIDQWNAKNSDITVGOYGG 120
DB 61 LVTRVVTTHMAHAGYNGADVGGADNSTIELTQNGFRNNATIDQWNAKNSDITVGOYGG 120
QY 121 NNAALVNQATSDSSVMVRQVGFNNATANQY 151
DB 121 NNAALVNQATSDSSVMVRQVGFNNATANQY 151
RESULT 9
```



CC the exemplification of the present invention  
XX Sequence 151 AA;  
SQ  
Query Match 75.8%; Score 574; DB 3; Length 151;  
Best Local Similarity 80.8%; Pred. No. 3.9e-49;  
Matches 122; Conservative 6; Mismatches 23; Indels 0; Gaps 0;  
QY 1 MKLLKVAFAAIVVSGSALAGVYDQVTRVVTHEMAHAGSDPSTLSIYQGSANAALALQ 60  
DB 1 MKLLKVAFAAIVVSGSALAGVYDQVTRVVTHEMAHAGSDPSTLSIYQGSANAALALQ 60  
QY 61 SDARKSETTITQSGYNGADVGQADNSTIELTONGFRNNATIDOWNAKNSDITVGOYGG 120  
DB 61 SDARKSETTITQSGYNGADVGQADNSTIELTONGFRNNATIDOWNAKNSDITVGOYGG 120  
QY 121 NNAALVNQTSASSVVMVRQVGFNNATANQY 151  
DB 121 NNAALVNQTSASSVVMVRQVGFNNATANQY 151  
RESULT 11  
AAB36354  
ID AAB36354 standard; protein; 151 AA.  
XX AAB36354;  
AC AAB36354;  
XX 26-FEB-2001 (first entry)  
DT Agfa::PT3#9 amino acid sequence SEQ ID NO:28.  
DE Salmonella; agfa; chromosomal gene replacement; fimbrin; epitope;  
KW vaccine; immune response; immunogen.  
XX Salmonella enteritidis.  
OS Escherichia coli.  
OS Synthetic.  
XX WO2000060102-A2.  
PN 12-OCT-2000.  
PD 05-APR-2000; 2000WO-CA000356.  
PF 05-APR-1999; 99US-0127888P.  
PR (UYVI-) UNIV VICTORIA.  
PA White AP, Doran JL, Collison SK, Kay WW;  
XX WPI; 2000-672631/65.  
XX N-PSDB; AAC64630.  
XX Recombinant agfa gene having a segment replaced by a foreign DNA sequence  
PT which encodes foreign epitope or antigen, expresses recombinant Agfa  
PT protein useful for eliciting immune response in animal.  
XX Disclosure; Page 138; 139pp; English.  
PS  
XX The present invention describes a recombinant agfa gene (I) where a  
CC segment of the gene has been replaced by a segment of a foreign DNA  
CC sequence which encodes a foreign epitope or antigen. Also described are:  
CC (1) use of thin aggregative fimbriae (SEF17/TAF) nucleation depended  
CC assembly system of strains of Salmonella, Escherichia coli and  
CC Enterobacteriaceae for the production of fimbriae comprising recombinant  
CC Agfa, CsgA and Agfa-homologue fimbrial subunits, respectively; (2)  
CC directing recombination of a recombinant gene into the chromosome of the  
CC homologous species; (3) directing recombination of a recombinant gene  
CC back into the chromosome of the homologous species, replacing the native  
CC copy of that gene; and (4) eliciting an immune response in an animal,  
CC comprising separating an amino acid polymer comprising a recombinant Agfa  
CC protein containing a replacement segment or segments of foreign amino  
CC acid sequence or sequences grown on a Salmonella, E. coli or

CC Enterobacteriaceae host cell, from the host cell and introducing the  
CC polymer into the animal in conjunction with a carrier or diluent. (I) is  
CC useful for the expression of recombinant Agfa protein which is useful for  
CC eliciting an immune response in an animal. In a fimbrial presentation  
CC system the heterologous antigens are presented in high numbers (up to  
CC 500,000 copies/cell), the hybrid fimbrial protein possesses both the  
CC immunogenicity and adhesion properties relevant for an efficient live  
CC vaccine, the carrier fimbrial subunit proteins are usually strong  
CC immunogens, which may be important for directing an immune response  
CC against the inserted epitope, and hybrid fimbriae are easy and  
CC inexpensive to purify in large amount. The present sequence is given in  
CC the exemplification of the present invention  
XX Sequence 151 AA;  
SQ  
Query Match 75.0%; Score 568; DB 3; Length 151;  
Best Local Similarity 80.1%; Pred. No. 1.5e-48;  
Matches 121; Conservative 4; Mismatches 26; Indels 0; Gaps 0;  
QY 1 MKLLKVAFAAIVVSGSALAGVYDQVTRVVTHEMAHAGSDPSTLSIYQGSANAALALQ 60  
DB 1 MKLLKVAFAAIVVSGSALAGVYDQVTRVVTHEMAHAGSDPSTLSIYQGSANAALALQ 60  
QY 61 SDARKSETTITQSGYNGADVGQADNSTIELTONGFRNNATIDOWNAKNSDITVGOYGG 120  
DB 61 SDARKSETTITQSGYNGADVGQADNSTIELTONGFRNNATIDOWNAKNSDITVGOYGG 120  
QY 121 NNAALVNQTSASSVVMVRQVGFNNATANQY 151  
DB 121 NNAALVNQTSASSVVMVRQVGFNNATANQY 151  
RESULT 12  
AAB36351  
ID AAB36351 standard; protein; 151 AA.  
XX AAB36351;  
AC AAB36351;  
XX 26-FEB-2001 (first entry)  
DT Agfa::PT3#6 amino acid sequence SEQ ID NO:22.  
DE Salmonella; agfa; chromosomal gene replacement; fimbrin; epitope;  
KW vaccine; immune response; immunogen.  
XX Salmonella enteritidis.  
OS Escherichia coli.  
OS Synthetic.  
XX WO2000060102-A2.  
PN 12-OCT-2000.  
PD 05-APR-2000; 2000WO-CA000356.  
PF 05-APR-1999; 99US-0127888P.  
PR (UYVI-) UNIV VICTORIA.  
PA White AP, Doran JL, Collison SK, Kay WW;  
XX WPI; 2000-672631/65.  
XX N-PSDB; AAC64627.  
XX Recombinant agfa gene having a segment replaced by a foreign DNA sequence  
PT which encodes foreign epitope or antigen, expresses recombinant Agfa  
PT protein useful for eliciting immune response in animal.  
XX Disclosure; Page 137; 139pp; English.  
PS  
XX The present invention describes a recombinant agfa gene (I) where a  
CC segment of the gene has been replaced by a segment of a foreign DNA  
CC sequence which encodes a foreign epitope or antigen. Also described are:

CC (1) use of thin aggregative fimbriae (SEF17/TAF) nucleation depended  
 CC assembly system of strains of Salmonella, Escherichia coli and  
 CC Enterobacteriaceae for the production of fimbriae comprising recombinant  
 CC Agfa, CsgA and Agfa-homologue fimbrial subunits, respectively; (2)  
 CC directing recombination of a recombinant gene into the chromosome of the  
 CC homologous species; (3) directing recombination of a recombinant gene  
 CC back into the chromosome of the homologous species, replacing the native  
 CC copy of that gene; and (4) eliciting an immune response in an animal,  
 CC comprising separating an amino acid polymer comprising a recombinant Agfa  
 CC protein containing a replacement segment or segments of foreign amino  
 CC acid sequence or sequences grown on a Salmonella, E. coli or  
 CC Enterobacteriaceae host cell, from the host cell and introducing the  
 CC polymer into the animal in conjunction with a carrier or diluent. (1) is  
 CC useful for the expression of recombinant Agfa protein which is useful for  
 CC eliciting an immune response in an animal. In a fimbrial presentation  
 CC system the heterologous antigens are presented in high numbers (up to  
 CC 500,000 copies/cell), the hybrid fimbrial protein possesses both the  
 CC immunogenicity and adhesion properties relevant for an efficient live  
 CC vaccine, the carrier fimbrial subunit proteins are usually strong  
 CC immunogens, which may be important for directing an immune response  
 CC against the inserted epitope, and hybrid fimbriae are easy and  
 CC inexpensive to purify in large amount. The present sequence is given in  
 CC the exemplification of the present invention

XX SQ Sequence 151 AA;

Query Match 74.9%; Score 567; DB 3; Length 151;  
 Best Local Similarity 80.8%; Pred. No. 1.9e-48;  
 Matches 122; Conservative 5; Mismatches 24; Indels 0; Gaps 0;  
 QY 1 MLLKVAFAAIVVSGSALAGVVDQLVTRVVTTHMAHAGSPDSTLSIYQYGSANAALALQ 60  
 DB 1 MLLKVAFAAIVVSGSALAGVVDQLVTRVVTTHMAHAGSPDSTLSIYQYGSANAALALQ 60  
 QY 61 SDARKSETTITQSGYNGADVGQADNSTIELTQNGFRNNATIDQWNAKNSDITVQYGG 120  
 DB 61 SDARKSETTITQSGYNGADVGQADNSTIELTQNGFRNNATIDQWNAKNSDITVQYGG 120

QY 121 NNAALVNQTASDSSVMVRQVGFNNATANQY 151

DB 121 NNAALVNQTASDSSVMVRQVGFNNATANQY 151

# RESULT 13

AAB36355  
 ID AAB36355 standard; protein; 151 AA.

XX AC AAB36355;

XX 26-FEB-2001 (first entry)

DE Agfa::PT3#10 amino acid sequence SEQ ID NO:30.

XX Salmonella; agfa; chromosomal gene replacement; fimbrial; epitope;  
 KW vaccine; immune response; immunogen.

XX Salmonella enteritidis.

OS Escherichia coli.

OS Synthetic.

XX WO2000060102-A2.

XX 12-OCT-2000.

XX 05-APR-2000; 2000WO-CA000356.

XX 05-APR-1999; 99US-0127888P.

XX (UYVI-) UNIV VICTORIA.

XX White AP, Doran JL, Collison SK, Kay WW;

DR WPI; 2000-672631/65.

DR N-PSDB; AAC64631.

XX Recombinant agfa gene having a segment replaced by a foreign DNA sequence  
 PT which encodes foreign epitope or antigen, expresses recombinant Agfa  
 PT protein useful for eliciting immune response in animal.

XX Disclosure; Page 139; 139pp; English.

XX The present invention describes a recombinant agfa gene (I) where a  
 CC segment of the gene has been replaced by a segment of a foreign DNA  
 CC sequence which encodes a foreign epitope or antigen. Also described are:  
 CC (1) use of thin aggregative fimbriae (SEF17/TAF) nucleation depended  
 CC assembly system of strains of Salmonella, Escherichia coli and  
 CC Enterobacteriaceae for the production of fimbriae comprising recombinant  
 CC Agfa, CsgA and Agfa-homologue fimbrial subunits, respectively; (2)  
 CC directing recombination of a recombinant gene into the chromosome of the  
 CC homologous species; (3) directing recombination of a recombinant gene  
 CC back into the chromosome of the homologous species, replacing the native  
 CC copy of that gene; and (4) eliciting an immune response in an animal,  
 CC comprising separating an amino acid polymer comprising a recombinant Agfa  
 CC protein containing a replacement segment or segments of foreign amino  
 CC acid sequence or sequences grown on a Salmonella, E. coli or  
 CC Enterobacteriaceae host cell, from the host cell and introducing the  
 CC polymer into the animal in conjunction with a carrier or diluent. (1) is  
 CC useful for the expression of recombinant Agfa protein which is useful for  
 CC eliciting an immune response in an animal. In a fimbrial presentation  
 CC system the heterologous antigens are presented in high numbers (up to  
 CC 500,000 copies/cell), the hybrid fimbrial protein possesses both the  
 CC immunogenicity and adhesion properties relevant for an efficient live  
 CC vaccine, the carrier fimbrial subunit proteins are usually strong  
 CC immunogens, which may be important for directing an immune response  
 CC against the inserted epitope, and hybrid fimbriae are easy and  
 CC inexpensive to purify in large amount. The present sequence is given in  
 CC the exemplification of the present invention

XX SQ Sequence 151 AA;

Query Match 74.8%; Score 566; DB 3; Length 151;

Best Local Similarity 80.8%; Pred. No. 2.4e-48;

Matches 122; Conservative 5; Mismatches 24; Indels 0; Gaps 0;

QY 1 MLLKVAFAAIVVSGSALAGVVDQLVTRVVTTHMAHAGSPDSTLSIYQYGSANAALALQ 60

DB 1 MLLKVAFAAIVVSGSALAGVVDQLVTRVVTTHMAHAGSPDSTLSIYQYGSANAALALQ 60

QY 61 SDARKSETTITQSGYNGADVGQADNSTIELTQNGFRNNATIDQWNAKNSDITVQYGG 120

DB 61 SDARKSETTITQSGYNGADVGQADNSTIELTQNGFRNNATIDQWNAKNSDITVQYGG 120

QY 121 NNAALVNQTASDSSVMVRQVGFNNATANQY 151

DB 121 NNAALVNQTASDSSVMVRQVGFNNATANQY 151

# RESULT 14

AAR62761

ID AAR62761 standard; protein; 120 AA.

XX AC AAR62761;

XX 27-AUG-2003 (revised)

DT 25-MAR-2003 (revised)

DT 26-JUN-1995 (first entry)

XX Agfa sequence.

XX Salmonella; Agfa; vaccine.

XX Salmonella enteritidis.

XX WO9425598-A2.

PD 10-NOV-1994.

XX PF 26-APR-1994; 94WO-IB000207.  
XX PR 26-APR-1993; 93US-00054452.  
XX PA (UYVI-) UNIV VICTORIA INNOVATION & DEV CORP.  
XX PA (KING/) KING J.  
XX PI Kay WW, Collinson SK, Clouthier SC, Doran JL;  
XX DR WPI; 1994-358275/44.  
XX DR N-PSDB; AAQ73066.  
XX PT Eliciting an immune response to Salmonella - using attenuated Salmonella  
XX PT strains, vector constructs, or compsns. contg. fimbrial type proteins.  
XX PS Disclosure; Fig 7A; 95pp; English.  
XX CC The sequence represents the Salmonella enteritidis 27655-3b TnpHoA mutant  
XX CC strain Agfa protein. The encoding DNA and isolated Agfa protein are used  
XX CC in genetic immunization and vaccine compositions, respectively, to elicit  
XX CC an immune response to Salmonella in animals (e.g. food producing animals)  
XX CC and humans. (Updated on 25-MAR-2003 to correct PN field.) (Updated on 27-  
XX CC AUG-2003 to correct OS field.)  
XX SQ Sequence 120 AA;  
Query Match 64.3%; Score 487; DB 2; Length 120;  
Best Local Similarity 97.9%; Pred. No. 1.3e-40;  
Matches 95; Conservative 2; Mismatches 0; Indels 0; Gaps 0;  
QY 37 HASGPDSTLSIYQYGSANAALQSDARKSETTITQSGYNGADVGGADNSTIELTQNG 96  
Db 16 NSSGPDSTLSIYQYGSANAALQSDARKSETTITQSGYNGADVGGADNSTIELTQNG 75  
QY 97 FRNATIDQWNAKNSDITVQYGGNNAALVNQTASDS 133  
Db 76 FRNATIDQWNAKNSDITVQYGGNNAALVNQTASDS 112  
RESULT 15  
AAW23569  
ID AAW23569 standard; protein; 120 AA.  
XX AC AAW23569;  
XX DT 25-MAR-2003 (revised)  
XX DT 29-SEP-1997 (first entry)  
XX DE Salmonella enteritidis 27655-3b TnpHoA mutant agfa fragment.  
XX KW Enteropathogenic bacteria; enterobacteria; S.enteritidis; antibody.  
XX OS Salmonella enteritidis.  
XX PN US5635617-A.  
XX PD 03-JUN-1997.  
XX PF 26-APR-1994; 94US-00233788.  
XX PR 26-APR-1993; 93US-00054452.  
XX PA (UYVI-) UNIV VICTORIA INNOVATION & DEV CORP.  
XX PI Collinson SK, Kay WW, Doran JL;  
XX DR WPI; 1997-309886/28.  
XX DR N-ESDB; AAT74141.  
XX PT Isolated Salmonella gene agfa - used for diagnosis of Salmonella or  
XX PT enteropathogenic bacteria of the Enterobacteria family.

PS Example 2; Fig 7; 85pp; English.

XX The present sequence represents an agfa fragment encoded by an agfa gene  
XX fragment derived from Salmonella enteritidis 27655-3b TnpHoA mutant  
XX strain. The nucleic acid can be used to provide diagnostic assays for  
XX Salmonella and/or enteropathogenic bacteria of the family Enterobacteria.  
XX It can also be used to provide proteins and antibodies which can be used  
XX for assays. The nucleic acid sequence can be used to provide probes or  
XX primers which can specifically hybridise to nucleic acid molecules from  
XX greater than 99% of Salmonella strains that are pathogenic to warm-  
XX blooded animals relative to nucleic acid molecules from virtually all  
XX other microbial organisms. (Updated on 25-MAR-2003 to correct PF field.)

SQ Sequence 120 AA;

Query Match 64.3%; Score 487; DB 2; Length 120;  
Best Local Similarity 97.9%; Pred. No. 1.3e-40;  
Matches 95; Conservative 2; Mismatches 0; Indels 0; Gaps 0;  
QY 37 HASGPDSTLSIYQYGSANAALQSDARKSETTITQSGYNGADVGGADNSTIELTQNG 96  
Db 16 NSSGPDSTLSIYQYGSANAALQSDARKSETTITQSGYNGADVGGADNSTIELTQNG 75  
QY 97 FRNATIDQWNAKNSDITVQYGGNNAALVNQTASDS 133  
Db 76 FRNATIDQWNAKNSDITVQYGGNNAALVNQTASDS 112

Search completed: August 2, 2004, 14:48:25  
Job time : 45.9 secs





GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: August 2, 2004, 14:40:48 ; Search time 12 seconds  
(without alignments)  
649.627 Million cell updates/sec

Title: US-09-543-407-16  
Perfect score: 757  
Sequence: 1 MKLLKVAFAAIVVSGSALA.....DSSVMVRQVFGNNATANCY 151

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 389414 seqs, 51625971 residues

Total number of hits satisfying chosen parameters: 389414

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Issued Patents AA.\*  
1: /cgn2\_6/ptodata/2/iaa/5A COMB.pcp.\*  
2: /cgn2\_6/ptodata/2/iaa/5B COMB.pcp.\*  
3: /cgn2\_6/ptodata/2/iaa/6A COMB.pcp.\*  
4: /cgn2\_6/ptodata/2/iaa/6B COMB.pcp.\*  
5: /cgn2\_6/ptodata/2/iaa/PCFUS COMB.pcp.\*  
6: /cgn2\_6/ptodata/2/iaa/BACKFILES.pcp.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	653	86.3	151	1	US-08-233-788A-59
2	487	64.3	120	1	US-08-233-788A-57
3	88	11.6	2315	4	US-09-543-681A-5434
4	86.5	11.4	943	4	US-09-477-135A-131
5	85	11.2	1415	4	US-09-252-991A-26438
6	82	10.8	1026	1	US-08-194-290-7
7	82	10.8	1026	2	US-08-614-377A-7
8	82	10.8	1026	3	US-09-142-648B-7
9	81.5	10.8	746	4	US-09-134-001C-3214
10	81	10.7	690	4	US-08-336-115C-12
11	81	10.7	691	4	US-08-336-115C-6
12	81	10.7	745	4	US-09-346-115C-6
13	80.5	10.6	1095	4	US-09-206-942-69
14	80.5	10.6	1536	1	US-08-038-682-2
15	80.5	10.6	1536	1	US-08-302-832-2
16	80.5	10.6	1536	2	US-08-530-198-2
17	80.5	10.6	1536	2	US-08-469-880-2
18	80.5	10.6	1536	2	US-08-728-470-2
19	80.5	10.6	1536	2	US-08-617-697-2
20	80.5	10.6	1536	3	US-08-719-641-2
21	80.5	10.6	1536	4	US-09-206-942-67
22	79.5	10.5	720	1	US-07-731-157A-2
23	79.5	10.5	720	2	US-08-541-780-2
24	79.5	10.5	1739	4	US-09-540-236-3739
25	79	10.4	907	1	US-07-718-575-2
26	79	10.4	907	1	US-08-481-206-2
27	79	10.4	907	2	US-08-486-269A-2

28	78	10.3	1864	2	US-08-804-227C-3	Sequence 3, Appli
29	77.5	10.2	518	3	US-09-043-123-2	Sequence 2, Appli
30	77.5	10.2	2314	4	US-09-268-347-49	Sequence 49, Appl
31	77	10.2	339	4	US-09-252-991A-32096	Sequence 32096, A
32	77	10.2	1612	1	US-08-169-927-2	Sequence 2, Appli
33	77	10.2	10182	4	US-09-134-001C-3159	Sequence 3159, Ap
34	76.5	10.1	892	4	US-09-336-447A-5	Sequence 5, Appli
35	76	10.0	273	4	US-09-328-352-6167	Sequence 6167, Ap
36	76	10.0	741	4	US-09-489-039A-7854	Sequence 7854, Ap
37	76	10.0	906	1	US-08-254-573-2	Sequence 2, Appli
38	76	10.0	906	1	US-08-687-379-2	Sequence 2, Appli
39	76	10.0	906	1	US-08-687-379-4	Sequence 4, Appli
40	76	10.0	906	4	US-08-172-332-1	Sequence 1, Appli
41	76	10.0	906	4	US-08-216-326-2	Sequence 2, Appli
42	76	10.0	943	4	US-09-056-556-204	Sequence 204, App
43	76	10.0	943	4	US-09-072-596-199	Sequence 199, App
44	76	10.0	943	4	US-09-072-567-204	Sequence 204, App
45	75	9.9	359	4	US-09-328-352-6562	Sequence 6562, Ap

ALIGNMENTS

RESULT 1  
US-08-233-788A-59  
; Sequence 59, Application US/08233788A  
; Patent No. 5635617  
; GENERAL INFORMATION:  
; APPLICANT: Doran, James L.  
; APPLICANT: Kay, William W.  
; APPLICANT: Collinson, Karen S.  
; APPLICANT: Clouthier, Sharon C.  
; TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR DETECTION  
; TITLE OF INVENTION: OF SALMONELLA  
; NUMBER OF SEQUENCES: 61  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Seed and Berry  
; STREET: 6300 Columbia Center, 701 Fifth Avenue  
; CITY: Seattle  
; STATE: Washington  
; COUNTRY: U.S.A.  
; ZIP: 98104-7892  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25  
; CURRENT APPLICATION NUMBER: US/08/233,788A  
; FILING DATE: 26-APR-1994  
; CLASSIFICATION: 435  
; ATTORNEY/AGENT INFORMATION:  
; NAME: King, Joshua  
; REGISTRATION NUMBER: 35,570  
; REFERENCE/DOCKET NUMBER: 920043.403C2  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (206) 622-4900  
; TELEFAX: (206) 682-6031  
; TELEX: 3723836 SEEDANBERRY  
; INFORMATION FOR SEQ ID NO: 59:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 151 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
US-08-233-788A-59

Query Match 86.3%; Score 653; DB 1; Length 151;  
Best Local Similarity 89.4%; Pred. No. 1.8e-60;  
Matches 135; Conservative 3; Mismatches 13; Indels 0; Gaps 0;

1 MKLLKVAFAAIVVSGSALAGVYDQVTRVVVTHMAHAGSPDSTLSTYQVGSNAALAQ 60

Ds 1 MLLKVAFAAIVVSGSALAGVVPQMGQGNHGGNGSGPDSLTLSIYQVGSANAALALQ 60  
Qy 61 SDARKSETTITQSGYNGADVGQADNSTIETLQNGFRNNATIDQWAKNSDITVQYGG 120  
Ds 61 SDARKSETTITQSGYNGADVGQADNSTIETLQNGFRNNATIDQWAKNSDITVQYGG 120  
Qy 121 NNAALVNOTASDSSVMVRQVGFNNATANY 151  
Ds 121 NNAALVNOTASDSSVMVRQVGFNNATANY 151

## RESULT 2

US-08-233-788A-57  
; Sequence 57, Application US/08233788A  
; Patent No. 5635617  
; GENERAL INFORMATION:  
; APPLICANT: Doran, James L.  
; APPLICANT: Kay, William W.  
; APPLICANT: Collinson, Karen S.  
; APPLICANT: Clothier, Sharon C.  
; TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR DETECTION  
; TITLE OF INVENTION: OF SALMONELLA  
; NUMBER OF SEQUENCES: 61  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Seed and Berry  
; STREET: 6300 Columbia Center, 701 Fifth Avenue  
; CITY: Seattle  
; STATE: Washington  
; COUNTRY: U.S.A.  
; ZIP: 98104-7092  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patent in Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/233,788A  
; FILING DATE: 26-APR-1994  
; CLASSIFICATION: 435  
; ATTORNEY/AGENT INFORMATION:  
; NAME: King, Joshua  
; REGISTRATION NUMBER: 35,570  
; REFERENCE/DOCKET NUMBER: 920043.403C2  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (206) 622-4900  
; TELEFAX: (206) 682-6031  
; TELEX: 3723836 SEEDANBERRY  
; INFORMATION FOR SEQ ID NO: 57:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 120 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
US-08-233-788A-57

Query Match 64.3%; Score 487; DB 1; Length 120;  
Best Local Similarity 97.9%; Pred. No. 2.5e-43;  
Matches 95; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Qy 37 HASGPDSTLSIYQVGSANAALALQSDARKSETTITQSGYNGADVGQADNSTIETLQNG 96  
Ds 16 NSSGPDSTLSIYQVGSANAALALQSDARKSETTITQSGYNGADVGQADNSTIETLQNG 75  
Qy 97 FRNNATIDQWAKNSDITVQYGGNNAALVNOTASDS 133  
Ds 76 FRNNATIDQWAKNSDITVQYGGNNAALVNOTASDS 112

## RESULT 3

US-09-543-681A-5434  
; Sequence 5434, Application US/09543681A  
; Patent No. 6605709  
; GENERAL INFORMATION:

; APPLICANT: GARY BRETON  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PROTEUS MIRABILIS  
; TITLE OF INVENTION: DIAGNOSTICS AND THERAPEUTICS  
; FILE REFERENCE: 2709.1002-001  
; CURRENT APPLICATION NUMBER: US/09/543,681A  
; CURRENT FILING DATE: 2000-04-05  
; PRIOR APPLICATION NUMBER: US 60/128,706  
; PRIOR FILING DATE: 1999-04-09  
; NUMBER OF SEQ ID NOS: 8344  
; SEQ ID NO 5434  
; LENGTH: 2315  
; TYPE: PRT  
; ORGANISM: Proteus mirabilis  
US-09-543-681A-5434

Query Match 11.6%; Score 88; DB 4; Length 2315;  
Best Local Similarity 28.0%; Pred. No. 6;  
Matches 37; Conservative 15; Mismatches 54; Indels 26; Gaps 6;  
Qy 24 DQVTRVVTHEMAHAGPD-STLSIYQVGSANAALALQSDARKSETTITQSGYNGADVG 82  
Ds 1700 DVTVTITLNDAGHASAPDYSTLAGSQH---DGKIALHGTGK---VTYDGAFTVTVI 1752  
Qy 83 QGADNS-----TIELTQNGFRNNATIDQWAKNSDITVQYGG--NNAALVNOTASDS 133  
Ds 1753 PAGSKSVSFIVDPTEANQNAF-----NAEGKVVATITGTSNNATVTDIVNNA 1803  
Qy 134 SVMVRQVGFNN 145  
Ds 1804 GASATGVYDGN 1815

## RESULT 4

US-09-477-135A-131  
; Sequence 131, Application US/09477135A  
; Patent No. 6572865  
; GENERAL INFORMATION:  
; APPLICANT: Nano, Francis  
; TITLE OF INVENTION: Mycobacterium Tuberculosis DNA Sequences Encoding  
; TITLE OF INVENTION: Immunostimulatory Peptides  
; FILE REFERENCE: 52888  
; CURRENT APPLICATION NUMBER: US/09/477,135A  
; CURRENT FILING DATE: 2000-01-03  
; PRIOR APPLICATION NUMBER: 08990823  
; PRIOR FILING DATE: 1997-12-15  
; PRIOR APPLICATION NUMBER: US 96/10375  
; PRIOR FILING DATE: 1996-06-14  
; PRIOR APPLICATION NUMBER: 60/000,254  
; PRIOR FILING DATE: 1995-06-15  
; NUMBER OF SEQ ID NOS: 169  
; SOFTWARE: Patent In Ver. 2.0  
; SEQ ID NO 131  
; LENGTH: 943  
; TYPE: PRT  
; ORGANISM: Mycobacterium tuberculosis  
US-09-477-135A-131

Query Match 11.4%; Score 86.5; DB 4; Length 943;  
Best Local Similarity 24.8%; Pred. No. 2.4;  
Matches 39; Conservative 22; Mismatches 73; Indels 23; Gaps 6;  
Qy 7 AAFALVWSG-----SALAGVTDQLVTRVVTHEMAHAGPDSTLS-IYQVGSANA 56  
Ds 96 ALFVSLVWSNLLQONAPATAEAAYEQWAOQVAAMFGYHAGASAAVSAITPFGQALPT 155  
Qy 57 L-----ALQSDARKSETTITQSGYNGADVGQADNSTIETLQNGFRNNATIDQWAKNSD 112  
Ds 156 VAGGGALVSAQAQVTRVTRFNLGL-ANVGEQN-----VGNVGNVFNLSANIGN 207  
Qy 113 ITVQYGGNNAALVNOTASDSVMVRQVGFNNAT 149  
Ds 208 IGSNGIGSSNIGFGN-VGFLTAALNNIGFGTGSNN 243

RESULT 5  
US-09-252-991A-26438  
; Sequence 26438, Application US/09252991A  
; Patent No. 6551795  
; GENERAL INFORMATION:  
; APPLICANT: Marc J. Rubenfield et al.  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS  
; FILE REFERENCE: 107196.136  
; CURRENT APPLICATION NUMBER: US/09/252.991A  
; PRIOR FILING DATE: 1999-02-18  
; PRIOR APPLICATION NUMBER: US 60/074,788  
; PRIOR FILING DATE: 1998-02-18  
; PRIOR APPLICATION NUMBER: US 60/094,190  
; PRIOR FILING DATE: 1998-07-27  
; NUMBER OF SEQ ID NOS: 33142  
; SEQ ID NO 26438  
; LENGTH: 1415  
; TYPE: PRT  
; ORGANISM: Pseudomonas aeruginosa  
US-09-252-991A-26438

Query Match 11.2%; Score 85; DB 4; Length 1415;  
Best Local Similarity 25.6%; Pred. No. 6.1;  
Matches 43; Conservative 19; Mismatches 54; Indels 52; Gaps 7;  
QY 24 DQLVVTHM-----AHASGPDSTLSIYQGSANA-----ALA 58  
Db 357 DQAGTWRILSHVVAAGVAGGQGVHVAQGPAGANASNGVTIYQQQPAVD 416  
QY 59 LQDARKSETTITQSYG-----NGADVGGADNSTIETLQN-----GFR 98  
Db 417 LAAGANGTSVQSGANGISVQSQNGANGICAGASDTSVVSQNSPNIGSGVN 476  
QY 99 NNATIDOWNAX-----SDITVQYGGNNALVNTASDSSVMVRQVG 141  
Db 477 GVTIVQSQNGANGISGASGITVQ--SONGANIGSGASG:SVVSQSG 522

RESULT 6  
US-08-194-290-7  
; Sequence 7, Application US/08194290  
; Patent No. 5500353  
; GENERAL INFORMATION:  
; APPLICANT: Smit, John  
; APPLICANT: Bingle, Wade H  
; TITLE OF INVENTION: Bacterial surface protein expression  
; NUMBER OF SEQUENCES: 10  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Shlesinger, Arkwright & Garvey  
; STREET: 3000 South Eads Street  
; CITY: Arlington  
; STATE: Virginia  
; COUNTRY: USA  
; ZIP: 22202  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patent In Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/194,290  
; FILING DATE: 09-FEB-1994  
; CLASSIFICATION: 435  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Garvey, George A  
; REGISTRATION NUMBER: 17737  
; REFERENCE/DOCKET NUMBER: 5946-1  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 703-684-5600  
; TELEFAX: 703-836-5288  
; INFORMATION FOR SEQ ID NO: 7:

; SEQUENCE CHARACTERISTICS:  
; LENGTH: 1026 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
US-08-194-290-7  
Query Match 10.8%; Score 82; DB 1; Length 1026;  
Best Local Similarity 24.8%; Pred. No. 7.9;  
Matches 34; Conservative 28; Mismatches 55; Indels 20; Gaps 7;  
QY 11 AIVVSGSALAGYDOLVTRVVTHEMAHA---SGPDSLSIYQGSANAALALQSDARKSE 67  
Db 463 AAATAGATVAGRVNGAVT--ITDAAASATTAGTATVTLGSFGAA---TIDSSAL--- 513  
QY 68 TTTITQSGYNGADVQGA-----DNSTIELQNGFRNNATI-DOWNAKNSDITVQYQGN 121  
Db 514 TTVNLSGTGSLIGRGALATPATNTLTNVNGLTTTGATDSEAAADGFTTINIAGS 573  
QY 122 NAA--LVNQTSADSSVM 136  
Db 574 TASSTIASLVAADATTL 590

RESULT 7  
US-08-614-377A-7  
; Sequence 7, Application US/08614377A  
; Patent No. 5976864  
; GENERAL INFORMATION:  
; APPLICANT: Smit, John  
; APPLICANT: Bingle, Wade H  
; APPLICANT: No. 5976864ellini, John F.  
; TITLE OF INVENTION: EXPRESSION AND SECRETION OF  
; TITLE OF INVENTION: HETEROLOGOUS  
; TITLE OF INVENTION: POLYPEPTIDES FROM CAULOBACTER  
; NUMBER OF SEQUENCES: 12  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Fish & Richardson PC  
; STREET: 225 Franklin Street  
; CITY: Boston  
; STATE: Massachusetts  
; COUNTRY: USA  
; ZIP: 02110-2804  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patent In Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/614,377A  
; FILING DATE: 12-MAR-1996  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/194,290  
; FILING DATE: 09-FEB-1994  
; CLASSIFICATION: 435  
; APPLICATION NUMBER: US 07/895,367  
; FILING DATE: 09-JUNE-1992  
; CLASSIFICATION: 435  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Tsao, Y. Rocky  
; REGISTRATION NUMBER: 34053  
; REFERENCE/DOCKET NUMBER: 08106/002001  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 617-542-5070  
; TELEFAX: 617-542-8906  
; INFORMATION FOR SEQ ID NO: 7:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 1026 amino acids  
; TYPE: amino acid  
; STRANDEDNESS:  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein



```

; APPLICANT: Weltzin, Richard A.
; APPLICANT: Guy, Bruno
; TITLE OF INVENTION: LT and CT in Parenteral Immunization
; FILE REFERENCE: 06132/055002
; CURRENT APPLICATION NUMBER: US/09/336,115C
; CURRENT FILING DATE: 1999-06-18
; PRIOR APPLICATION NUMBER: US 09/100,258
; NUMBER OF SEQ ID NOS: 24
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 12
; LENGTH: 691
; TYPE: PRT
; ORGANISM: Helicobacter pylori
; FEATURE:
; NAME/KEY: SIGNAL
; LOCATION: (1)...(18)
US-09-336-115C-12

Query Match          10.7%; Score 81; DB 4; Length 691;
Best Local Similarity 23.1%; Pred. No. 5.7; 66; Indels 48; Gaps 5;
Matches 33; Conservative 27; Mismatches 18;

QY      8 AFAAIVSGSALAGVYDQLVTRV-----VTHEMAHAGSD 42
DB      100 AYAVALAALNAAGVMQVIALFIGCGPPTNQSYQSGFNTPALNGTITTCNQAYGTGPN 159

QY      43 STLSIYQYGSANAALQSDARKSETTITQSGYNGADVQCGADNSTIELTQNGFR----- 98
DB      160 GILSIDYQKLNQAYQI-----IQTALNQCGGMPALNDTTKTGVVNIQQTNYRTTQ 213

QY      99 NNATIDQNAKNSDITVQYGG-----NNAALVNOTASDSSVMVRQ 139
DB      214 NNIEHYTENGKEIPVSYSGSGSFPTIQITVHNRAENLLQQAATIMQVLITQ 267

RESULT 12
US-09-336-115C-6
; Sequence 6, Application US/09336115C
; Patent No. 6576244
; GENERAL INFORMATION:
; APPLICANT: Weltzin, Richard A.
; APPLICANT: Guy, Bruno
; TITLE OF INVENTION: LT and CT in Parenteral Immunization
; FILE REFERENCE: 06132/055002
; CURRENT APPLICATION NUMBER: US/09/336,115C
; CURRENT FILING DATE: 1999-06-18
; PRIOR APPLICATION NUMBER: US 09/100,258
; NUMBER OF SEQ ID NOS: 24
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 6
; LENGTH: 745
; TYPE: PRT
; ORGANISM: Helicobacter pylori
; FEATURE:
; NAME/KEY: SIGNAL
; LOCATION: (1)...(20)
; NAME/KEY: VARIANT
; LOCATION: 721
; OTHER INFORMATION: Xaa = Any Amino Acid
US-09-336-115C-6

Query Match          10.7%; Score 81; DB 4; Length 745;
Best Local Similarity 23.1%; Pred. No. 6.4; 68; Indels 4; Gaps 2;
Matches 27; Conservative 18; Mismatches 18;

QY      31 VTHEMAHAGSDTSLSIYQYGSANAALQSDARKSETTITQSGYNGADVQCGADNSTI 90
DB      152 ITCNRFSTGPGKMSIDFEEKLNQAYQIIQQALKNSGFPFLG-GNGTKV---SVNYYN 207

```

```

QY      91 ELTQGFNNATIDQNAKNSDITVQYGGNNAALVNOTASDSSVMVRQVQFGNNAT 147
DB      208 ECRQADINGVGYOFCARNGSSSSNGSGSTQTTATTIQQDGVTTITTTNNKAT 264

RESULT 13
US-09-206-942-69
; Sequence 69, Application US/09206942
; Patent No. 6432869
; GENERAL INFORMATION:
; APPLICANT: Loosmore, Sheena M.
; APPLICANT: Yang, Yan-Ping
; APPLICANT: Klein, Michel H.
; TITLE OF INVENTION: Protective Recombinant Haemophilus Influenzae High
; TITLE OF INVENTION: Molecular Weight Proteins
; FILE REFERENCE: 1038-861 MIS:jib
; CURRENT APPLICATION NUMBER: US/09/206,942
; EARLIER FILING DATE: 1998-12-08
; EARLIER APPLICATION NUMBER: 09/167,568
; EARLIER FILING DATE: 1998-10-07
; NUMBER OF SEQ ID NOS: 95
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 69
; LENGTH: 1095
; TYPE: PRT
; ORGANISM: Haemophilus influenzae
US-09-206-942-69

Query Match          10.6%; Score 80.5; DB 4; Length 1095;
Best Local Similarity 26.4%; Pred. No. 12; 60; Indels 27; Gaps 7;
Matches 39; Conservative 22; Mismatches 60;

QY      15 SGSALAGVYDQ--LVTRVVTH---EMAHAGPDPSTLSIYQYGSANAALQOS-DARKSET 68
DB      727 TGSILGGIESSGSVTLTATEGALAVNSIG--NVTVTANSALTTLAGSTIKGTESVT 784

QY      69 TITQSGYNGADVQCGADNSTIEL-----TONGFRNATIDQNAKNSDITVQYGGN 121
DB      785 TSSQSG-----DIGGTISGTVKATESLTQSNKIKATTEANVTSAUTGTTGTTISG 839

QY      122 NAALVNOTASDSSVMVRQVQFGNNATAN 149
DB      840 NTVNVTANAGDLTV-----GNGAEIN 860

RESULT 14
US-08-038-682-2
; Sequence 2, Application US/08038682
; Patent No. 5549897
; GENERAL INFORMATION:
; APPLICANT: BARENKAMP, STEPHEN J
; APPLICANT: ST. GEME III, JOSEPH W
; TITLE OF INVENTION: HIGH MOLECULAR WEIGHT SURFACE PROTEINS
; TITLE OF INVENTION: OF NON-TYPEABLE HAEMOPHILUS
; NUMBER OF SEQUENCES: 8
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Shoemaker and Mattare, Ltd
; STREET: 2001 Jefferson Davis Hwy., 1203 Crystal Plaza
; CITY: Arlington
; STATE: Virginia
; COUNTRY: U.S.A.
; ZIP: 22202-0286
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/038,682
; FILING DATE: 16-MAR-1993
; CLASSIFICATION: 424

```

ATTORNEY/AGENT INFORMATION:  
NAME: BERKSTRESSER, JERRY W  
REGISTRATION NUMBER: 22,651  
REFERENCE/DOCKET NUMBER: 1038-293  
TELEPHONE: (703) 415-0810  
TELEFAX: (703) 415-0813  
INFORMATION FOR SEQ ID NO: 2:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1536 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-038-682-2

Query Match 10.6%; Score 80.5; DB 1; Length 1536;  
Best Local Similarity 26.4%; Pred. No. 20;  
Matches 39; Conservative 22; Mismatches 60; Indels 27; Gaps 7;  
QY 15 SGSALAGVYDQ--LVTRVTH---EMAHASGPDSTLSIYQYGSANAALALQS-DARKSET 68  
DB 1168 TGSILGGIESSSGSVTLTATEGALAVSNISG--NTVTVTANSGLTTLGASTIKGTESVT 1225  
QY 69 TITQSGYNGADVGQGDNSTIEL-----TQNGFRNNATIDOWNAKNSDITVGOYGGN 121  
DB 1226 TSSQSG-----DIGGTISGGTVEVKATESLTQNSKIKATTGEANTVSATGTIGGTISG 1280  
QY 122 NAALVNQTASDSSVMVRQVGFNNATAN 149  
DB 1281 NTVNVTANAGDLTV-----GNGAEIN 1301

Query Match 10.6%; Score 80.5; DB 1; Length 1536;  
Best Local Similarity 26.4%; Pred. No. 20;  
Matches 39; Conservative 22; Mismatches 60; Indels 27; Gaps 7;  
QY 15 SGSALAGVYDQ--LVTRVTH---EMAHASGPDSTLSIYQYGSANAALALQS-DARKSET 68  
DB 1168 TGSILGGIESSSGSVTLTATEGALAVSNISG--NTVTVTANSGLTTLGASTIKGTESVT 1225  
QY 69 TITQSGYNGADVGQGDNSTIEL-----TQNGFRNNATIDOWNAKNSDITVGOYGGN 121  
DB 1226 TSSQSG-----DIGGTISGGTVEVKATESLTQNSKIKATTGEANTVSATGTIGGTISG 1280  
QY 122 NAALVNQTASDSSVMVRQVGFNNATAN 149  
DB 1281 NTVNVTANAGDLTV-----GNGAEIN 1301

Search completed: August 2, 2004, 14:58:32  
Job time : 12 secs

RESULT 15  
US-08-302-832-2  
Sequence 2, Application US/08302832  
Patent No. 5603938  
GENERAL INFORMATION:  
APPLICANT: Barenkamp, Stephen J  
TITLE OF INVENTION: High Molecular Weight Surface Proteins  
TITLE OF INVENTION: of No. 5603938-Typeable Haemophilus  
NUMBER OF SEQUENCES: 8  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Shoemaker and Mattare, Ltd.  
STREET: 2001 Jefferson Davis Hwy., 1203 Crystal Plaza  
STREET: Bldg. 1  
CITY: Arlington  
STATE: Virginia  
COUNTRY: U.S.A.  
ZIP: 22202-0286  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/302,832  
FILING DATE: 16-SEP-1994  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: GB 9205704.1  
FILING DATE: 16-MAR-1992  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US pct/us93/02166  
FILING DATE: 16-MAR-1993  
ATTORNEY/AGENT INFORMATION:  
NAME: Berkstresser, Jerry W  
REGISTRATION NUMBER: 22,651  
REFERENCE/DOCKET NUMBER: 1038-404  
TELEPHONE: (703) 415-0810  
TELEFAX: (703) 415-0813  
INFORMATION FOR SEQ ID NO: 2:

Result No.	Score	Query Match	Length	DB	ID	Description
1	481	63.5	151	12	US-09-741-873B-4	Sequence 4, Appl
2	481	63.5	151	12	US-09-741-873B-4	Sequence 4, Appl
3	414	54.7	131	12	US-09-741-873B-2	Sequence 2, Appl
4	414	54.7	131	12	US-09-741-873B-2	Sequence 2, Appl
5	113.5	15.0	445	15	US-10-369-493-25638	Sequence 26638,
6	93	12.3	290	12	US-10-282-132A-72503	Sequence 72503,
7	89.5	11.8	1852	12	US-10-282-132A-49757	Sequence 49757,
8	89	11.8	1852	14	US-10-238-075-1549	Sequence 1549, A
9	88.5	11.7	597	9	US-09-793-306-146	Sequence 146, App
10	86.5	11.4	154	16	US-10-437-963-162284	Sequence 162284,
11	86.5	11.4	943	9	US-09-596-634-131	Sequence 131, App
12	86.5	11.4	943	10	US-09-597-182-131	Sequence 131, App
13	86.5	11.4	943	10	US-09-597-181-131	Sequence 131, App
14	86.5	11.4	3300	12	US-10-282-132A-64369	Sequence 64369,
15	85	11.2	1554	16	US-10-383-930-38	Sequence 38, App





Db 19 SGNSELTQYGGNSALALQTDARNLDTITQHGNGADVGQSDSSIDLTORGF 78  
QY 99 NNATIDQWAKNSDITVQYGGNNAALVNGTASDSSVMVROVFGNNATANQY 151  
Db 79 NSATLDQWNGKSEMTVKQFGGNGAANDQATASNSVNVTVQVFGNNATARQY 131  
RESULT 5  
US-10-369-493-20638  
; Sequence 20638, Application US/10369493  
; Publication No. US20030233675A1  
; GENERAL INFORMATION:  
; APPLICANT: Cao, Yongwei  
; APPLICANT: Hinkle, Gregory J.  
; APPLICANT: Slater, Steven C.  
; APPLICANT: Goldman, Barry S.  
; APPLICANT: Chen, Xianfeng  
; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF  
; FILE REFERENCE: 38-10(52052)B  
; CURRENT APPLICATION NUMBER: US/10/369,493  
; PRIOR FILING DATE: 2003-02-28  
; PRIOR APPLICATION NUMBER: US 60/360,039  
; PRIOR FILING DATE: 2002-02-21  
; NUMBER OF SEQ ID NOS: 47374  
; SEQ ID NO 20638  
; LENGTH: 445  
; TYPE: PRT  
; ORGANISM: Rhodospseudomonas palustris  
; FEATURE:  
; NAME/KEY: unsure  
; LOCATION: (1)..(445)  
; OTHER INFORMATION: unsure at all Xaa locations  
US-10-369-493-20638  
Query Match 15.0%; Score 113.5; DB 15; Length 445;  
Best Local Similarity 25.0%; Pred. No. 0.0013;  
Matches 46; Conservative 18; Mismatches 63; Indels 57; Gaps 5;  
QY 3 LLKVAAPAAIIVWGSALAG---VYDQLVTRVVTHEMAHAGSPDSTLSIYQYGSANAALA 58  
Db 5 LFFVTASVLVLSSSAFAADSNVY-----LNQTGNDQANITQSGNNGSVGA 52  
QY 59 -----LQSDARKSETTI-----TQSGYNGADVGQADNSTIELT 93  
Db 53 FNGNSGFLQENGLTSGANLLTVKQSGNSVGRDIQKQSGAGNSAAIFQEGTGSDELQ 112  
QY 94 QNGPRNNATIDQWAKN-----SDITVQYGGNNAALVNGTASDSSVMV 137  
Db 113 QTGTNGAVPSGWNWINDPGVFNKITQDSSNGSKUSVIQDGKNNVFSIKQNTGNSTSV 172  
QY 138 RQVG 141  
Db 173 NQIG 176  
RESULT 6  
US-10-282-122A-72503  
; Sequence 72503, Application US/10282122A  
; Publication No. US20040029129A1  
; GENERAL INFORMATION:  
; APPLICANT: Wang, Liangsu  
; APPLICANT: Zamudio, Carlos  
; APPLICANT: Malone, Cheryl  
; APPLICANT: Haselbeck, Robert  
; APPLICANT: Ohlsen, Kari  
; APPLICANT: Zyskind, Judith  
; APPLICANT: Wall, Daniel  
; APPLICANT: Trawick, John  
; APPLICANT: Carr, Grant  
; APPLICANT: Yamamoto, Robert  
; APPLICANT: Forsyth, R.  
; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF  
; FILE REFERENCE: ELITRA.034A  
; CURRENT APPLICATION NUMBER: US/10/282,122A  
; CURRENT FILING DATE: 2003-02-20  
; PRIOR APPLICATION NUMBER: 60/191,078

; APPLICANT: Xu, H.  
; TITLE OF INVENTION: Identification of Essential Genes in Microorganisms  
; FILE REFERENCE: ELITRA.034A  
; CURRENT APPLICATION NUMBER: US/10/282,122A  
; CURRENT FILING DATE: 2003-02-20  
; PRIOR APPLICATION NUMBER: 60/191,078  
; PRIOR FILING DATE: 2000-03-21  
; PRIOR APPLICATION NUMBER: 60/206,848  
; PRIOR FILING DATE: 2000-05-23  
; PRIOR APPLICATION NUMBER: 60/207,727  
; PRIOR FILING DATE: 2000-05-26  
; PRIOR APPLICATION NUMBER: 60/230,335  
; PRIOR FILING DATE: 2000-09-06  
; PRIOR APPLICATION NUMBER: 60/230,347  
; PRIOR FILING DATE: 2000-09-09  
; PRIOR APPLICATION NUMBER: 60/242,578  
; PRIOR FILING DATE: 2000-10-23  
; PRIOR APPLICATION NUMBER: 60/253,625  
; PRIOR FILING DATE: 2000-11-27  
; PRIOR APPLICATION NUMBER: 60/257,931  
; PRIOR FILING DATE: 2000-12-22  
; PRIOR APPLICATION NUMBER: 60/267,636  
; PRIOR FILING DATE: 2001-02-09  
; PRIOR APPLICATION NUMBER: 60/269,308  
; PRIOR FILING DATE: 2001-02-16  
; Remaining Prior Application data removed - See File Wrapper or PALM.  
; NUMBER OF SEQ ID NOS: 78614  
; SOFTWARE: Patentin version 3.1  
; SEQ ID NO 72503  
; LENGTH: 290  
; TYPE: PRT  
; ORGANISM: Streptococcus mutans  
US-10-282-122A-72503  
Query Match 12.3%; Score 93; DB 12; Length 290;  
Best Local Similarity 26.7%; Pred. No. 0.12;  
Matches 40; Conservative 23; Mismatches 63; Indels 24; Gaps 5;  
QY 2 KLLKVAAPAAIIVWGSALAGVYDQLVTRVVTHEMAHAGSPDSTLSIYQYGSANAALAL-- 59  
Db 3 KLKIK-----STAKSGFRAYVLDSTET-VRTAQEEHQLSSSTVALGRTLIAQLAANQ 56  
QY 60 QSDARKSETTITQSGYNGADVGQADNSTIELTQNGPRNNATIDQWAKNSDITVQY 119  
Db 57 KQDSKVTVKVIGDSSFGHIIISVADTKGN-----VKGIQNGADVDKKTASGEVIVGPFM 110  
QY 120 GNNALVNGTASDSSVMVROVFGNNATAN 149  
Db 111 GNG-----QFWITDYGTGNPYTSS 130  
RESULT 7  
US-10-282-122A-49757  
; Sequence 49757, Application US/10282122A  
; Publication No. US20040029129A1  
; GENERAL INFORMATION:  
; APPLICANT: Wang, Liangsu  
; APPLICANT: Zamudio, Carlos  
; APPLICANT: Malone, Cheryl  
; APPLICANT: Haselbeck, Robert  
; APPLICANT: Ohlsen, Kari  
; APPLICANT: Zyskind, Judith  
; APPLICANT: Wall, Daniel  
; APPLICANT: Trawick, John  
; APPLICANT: Carr, Grant  
; APPLICANT: Yamamoto, Robert  
; APPLICANT: Forsyth, R.  
; APPLICANT: Xu, H.  
; TITLE OF INVENTION: Identification of Essential Genes in Microorganisms  
; FILE REFERENCE: ELITRA.034A  
; CURRENT APPLICATION NUMBER: US/10/282,122A  
; CURRENT FILING DATE: 2003-02-20  
; PRIOR APPLICATION NUMBER: 60/191,078

; PRIOR FILING DATE: 2000-03-21  
 ; PRIOR APPLICATION NUMBER: 60/206,848  
 ; PRIOR FILING DATE: 2000-05-23  
 ; PRIOR APPLICATION NUMBER: 60/207,727  
 ; PRIOR FILING DATE: 2000-05-26  
 ; PRIOR APPLICATION NUMBER: 60/230,335  
 ; PRIOR FILING DATE: 2000-09-06  
 ; PRIOR APPLICATION NUMBER: 60/230,347  
 ; PRIOR FILING DATE: 2000-09-09  
 ; PRIOR APPLICATION NUMBER: 60/242,578  
 ; PRIOR FILING DATE: 2000-10-23  
 ; PRIOR APPLICATION NUMBER: 60/253,625  
 ; PRIOR FILING DATE: 2000-11-27  
 ; PRIOR APPLICATION NUMBER: 60/257,931  
 ; PRIOR FILING DATE: 2000-12-22  
 ; PRIOR APPLICATION NUMBER: 60/267,636  
 ; PRIOR FILING DATE: 2001-02-09  
 ; PRIOR APPLICATION NUMBER: 60/269,308  
 ; PRIOR FILING DATE: 2001-02-16  
 ; Remaining Prior Application data removed - See File Wrapper or PALM.  
 ; NUMBER OF SEQ ID NOS: 78614  
 ; SOFTWARE: PatentIn version 3.1  
 ; SEQ ID NO 49757  
 ; LENGTH: 1862  
 ; TYPE: PRT  
 ; ORGANISM: Burkholderia fungorum  
 US-10-282-122A-49757  
  
 Query Match 11.8%; Score 89.5; DB 12; Length 1862;  
 Best Local Similarity 26.8%; Pred. No. 3.7;  
 Matches 51; Conservative 16; Mismatches 66; Indels 57; Gaps 8;  
  
 QY 7 AAFAAI-----VVGSGALAGVDYLVTRVVTHMAHA--SGPDSLSIYQGSANA 55  
 DB 1278 AAASAITATANGDVTASAGKLAAGTTGVTALNDINVAIESNGDAVLNA-CQSSLNA 1336  
  
 QY 56 ALALQSDARKSETTITQSGYNGADVQGGADNS-----TIELTQNGFRNNATIDOWN--- 107  
 DB 1337 TGGINSGA---ELTIT-----TGDLISLGASTAVGDDVTILNAGNALLGTLVGQNGYI 1388  
  
 QY 108 AKNSDIT-----VQVGGNNAALVNQTASDSSVMVR 138  
 DB 1389 SAGQDVTGPGTAFTNAVLGAQDIALTGSLQANSVQATGGDGAALNNVTSITTLTLTA 1448  
  
 QY 139 QVGGNNATA 148  
 DB 1449 NGNAGNDAA 1458  
  
 RESULT 8  
 US-10-238-075-1549  
 ; Sequence 1549, Application US/10238075  
 ; Publication No. US20030148324A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: I.N.S.E.R.M.  
 ; TITLE OF INVENTION: Polynucleotides which are of nature B2/D+ A- and which are isolated  
 ; TITLE OF INVENTION: E.coli, and biological uses of these polynucleotides and of their  
 ; FILE REFERENCE: BLANDINE  
 ; CURRENT APPLICATION NUMBER: US/10/238,075  
 ; PRIOR FILING DATE: 2002-09-10  
 ; PRIOR APPLICATION NUMBER: 0003145  
 ; PRIOR FILING DATE: 2000-03-10  
 ; NUMBER OF SEQ ID NOS: 1576  
 ; SOFTWARE: PatentIn version 3.1  
 ; SEQ ID NO 1549  
 ; LENGTH: 182  
 ; TYPE: PRT  
 ; ORGANISM: Escherichia coli  
 US-10-238-075-1549  
  
 Query Match 11.8%; Score 89; DB 14; Length 182;  
 Best Local Similarity 24.5%; Pred. No. 0.17;  
 Matches 39; Conservative 29; Mismatches 71; Indels 20; Gaps 7;

QY 1 MKLLKVAFAAIVVSGSALAGVDYLVTRVVTHMAHA--SGPDSLSIYQGSANA 56  
 DB 3 IKVIALATFVSAVPAAGSAMA--YDGTITFTGKVAQCTVNTSKDLAVTL----PTVAT 56  
  
 QY 57 LALQSDARKSETTITQSGYNGADVQGGADNS-----TIELTQNGFRNNATIDOWN 108  
 DB 57 SSLKDNAAATSLTFFAIRLTCATGMSAQNVAKAYFPPSSNIDLATHNLKNTAT--PTKA 114  
  
 QY 109 KNSDITVQYGGNNAALVNQTASDSSVMVRQVGGNNAT 147  
 DB 115 DNVOQLLNSNGTSTILLGEADNGQDVQSETIGSGSAT 153

## RESULT 9

US-09-793-306-146  
 ; Sequence 146, Application US/09793306  
 ; Patent No. US20020098200A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Campos-Neto, Antonio  
 ; APPLICANT: Skeiky, Yasir  
 ; APPLICANT: Ovendale, Pamela  
 ; APPLICANT: Jen, Shyian  
 ; APPLICANT: Lodes, Michael  
 ; APPLICANT: Corixa Corporation  
 ; TITLE OF INVENTION: Compounds and Methods for Diagnosis and Immunotherapy  
 ; TITLE OF INVENTION: of Tuberculosis  
 ; FILE REFERENCE: 014058-008740US  
 ; CURRENT APPLICATION NUMBER: US/09/793,306  
 ; CURRENT FILING DATE: 2001-02-26  
 ; PRIOR APPLICATION NUMBER: US 60/185,037  
 ; PRIOR FILING DATE: 2000-02-25  
 ; PRIOR APPLICATION NUMBER: US 60/223,828  
 ; PRIOR FILING DATE: 2000-08-08  
 ; NUMBER OF SEQ ID NOS: 164  
 ; SOFTWARE: PatentIn ver. 2.1  
 ; SEQ ID NO 146  
 ; LENGTH: 597  
 ; TYPE: PRT  
 ; ORGANISM: Artificial Sequence  
 ; FEATURE:  
 ; OTHER INFORMATION: Description of Artificial Sequence: mTTC#3-His  
 US-09-793-306-146

Query Match 11.7%; Score 88.5; DB 9; Length 597;  
 Best Local Similarity 35.9%; Pred. No. 1;  
 Matches 28; Conservative 7; Mismatches 32; Indels 11; Gaps 4;

QY 74 GYNGADVQGGADNSITELTQNGFRNNATIDQ--WNAKNSDITVQYGGNNAALVNQTAS 131  
 DB 387 GFGNAGDINTGFGNAGD--TNTGFGNAGFFNMGIGNAGNEDMGVGGSGFNVGVGN--AG 442  
  
 QY 132 DSSVMVRQVGGNNATAN 149  
 DB 443 NQS-----VGFGNAGTLN 455

## RESULT 10

US-10-437-963-162284  
 ; Sequence 162284, Application US/10437963  
 ; Publication No. US20040123343A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: La Rosa, Thomas J.  
 ; APPLICANT: Kovalic, David K.  
 ; APPLICANT: Zhou, Yihua  
 ; APPLICANT: Cao, Yongwei  
 ; APPLICANT: Wu, Wei  
 ; APPLICANT: Boukharov, Andrey A.  
 ; APPLICANT: Barbazuk, Brad  
 ; APPLICANT: Li, Ping  
 ; TITLE OF INVENTION: Rice Nucleic Acid Molecules and Other Molecules Associated With  
 ; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement  
 ; FILE REFERENCE: 38-21(53221)B

```
; CURRENT APPLICATION NUMBER: US/10/437,963
; CURRENT FILING DATE: 2003-05-14
; NUMBER OF SEQ ID NOS: 204966
; SEQ ID NO 162284
; LENGTH: 154
; TYPE: PRT
; ORGANISM: Oryza sativa
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT4530_61390C.1.pep
US-10-437-963-162284

Query Match      11.4%; Score 86.5; DB 16; Length 154;
Best Local Similarity 23.5%; Pred. No. 0.26;
Matches 40; Conservative 21; Mismatches 56; Indels 53; Gaps 7;

QY 5 KVAAPAAIVVSSALAGVDLVTRVVTTHMAHAGSPDSTLSIYQYGSANALALQSDAR 64
   |||||
Db 5 KLAALGFVILLIGLAS-----AVRVERYSNAQSGTSGEGTGSVNGAGAGLG----- 53
   |||||
QY 65 KSETTITQSGY-----GNGADVGGADNSTIETQNGFRNNA-TI 103
   |||||
Db 54 -SGTSGRSYNGAHAGSGGGGGGGYQYGGSGSGSGSSQTSQNGYYGSSAT 112
   |||||
QY 104 DOWNAKNSDITV-----QYQYGNNAALVNOTASDSSVMVRQVGFN 144
   |||||
Db 113 NTWYQYANAGNGGNGGQYGGSGAGV-----GSGS-----GFGD 151

RESULT 11
US-09-996-634-131
; Sequence 131, Application US/09996634
; Publication No. US20020172684A1
; GENERAL INFORMATION:
; APPLICANT: Nano, Francis
; TITLE OF INVENTION: Mycobacterium Tuberculosis DNA Sequences Encoding
; TITLE OF INVENTION: immunostimulatory Peptides
; FILE REFERENCE: 61260
; CURRENT APPLICATION NUMBER: US/09/996,634
; CURRENT FILING DATE: 2001-11-28
; PRIOR APPLICATION NUMBER: 09/447,135
; PRIOR FILING DATE: 2000-01-03
; PRIOR APPLICATION NUMBER: 08/990,823
; PRIOR FILING DATE: 1997-12-15
; PRIOR APPLICATION NUMBER: US 96/10375
; PRIOR FILING DATE: 1996-06-14
; PRIOR APPLICATION NUMBER: 60/000,254
; PRIOR FILING DATE: 1995-06-15
; NUMBER OF SEQ ID NOS: 169
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 131
; LENGTH: 943
; TYPE: PRT
; ORGANISM: Mycobacterium tuberculosis
US-09-996-634-131

Query Match      11.4%; Score 86.5; DB 9; Length 943;
Best Local Similarity 24.8%; Pred. No. 3.1;
Matches 39; Conservative 22; Mismatches 73; Indels 23; Gaps 6;

QY 7 AAFPAAIIVSG-----SALAGVDLVTRVVTTHMAHAGSPDSTLS-IYQYGSANAA 56
   |||||
Db 96 ALFVSLVSNLGNAPALAAATEAAVEQWMAQDVAAAMFGHAGASAASALTFFGQALPT 155
   |||||
QY 57 L-----ALQSDARKSETTITQSGYNGADYGGADNSTIETQNGFRNNAIDQWNAKNSD 112
   |||||
Db 156 VAGGALVSAQAQVTRVFRNLGL-ANVGEGN-----VGNVGNFNLGSANIGN 207
   |||||
QY 113 ITVQYGGNNAALVNOTASDSSVMVRQVGFNNATAN 149
   |||||
Db 208 IGSNGISSNIGFCN-VGPCLTAALNNIGFGNTGSNN 243

RESULT 12
US-09-996-634-131
; Sequence 131, Application US/09996634
; Publication No. US20020172684A1
; GENERAL INFORMATION:
; APPLICANT: Nano, Francis
; TITLE OF INVENTION: Mycobacterium Tuberculosis DNA Sequences Encoding
; TITLE OF INVENTION: immunostimulatory Peptides
; FILE REFERENCE: 61260
; CURRENT APPLICATION NUMBER: US/09/997,181
; CURRENT FILING DATE: 2001-11-28
; PRIOR APPLICATION NUMBER: 09/447,135
; PRIOR FILING DATE: 2000-01-03
; PRIOR APPLICATION NUMBER: 08/990,823
; PRIOR FILING DATE: 1997-12-15
; PRIOR APPLICATION NUMBER: US 96/10375
; PRIOR FILING DATE: 1996-06-14
; PRIOR APPLICATION NUMBER: 60/000,254
; PRIOR FILING DATE: 1995-06-15
; NUMBER OF SEQ ID NOS: 169
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 131
; LENGTH: 943
; TYPE: PRT
; ORGANISM: Mycobacterium tuberculosis
US-09-997-181-131

Query Match      11.4%; Score 86.5; DB 10; Length 943;
Best Local Similarity 24.8%; Pred. No. 3.1;
Matches 39; Conservative 22; Mismatches 73; Indels 23; Gaps 6;
```

```
US-09-997-182-131
; Sequence 131, Application US/09997182
; Publication No. US20030049263A1
; GENERAL INFORMATION:
; APPLICANT: Nano, Francis
; TITLE OF INVENTION: Mycobacterium Tuberculosis DNA Sequences Encoding
; TITLE OF INVENTION: immunostimulatory Peptides
; FILE REFERENCE: 61258
; CURRENT APPLICATION NUMBER: US/09/997,182
; CURRENT FILING DATE: 2001-11-28
; PRIOR APPLICATION NUMBER: 09/447,135
; PRIOR FILING DATE: 2000-01-03
; PRIOR APPLICATION NUMBER: 08/990,823
; PRIOR FILING DATE: 1997-12-15
; PRIOR APPLICATION NUMBER: US 96/10375
; PRIOR FILING DATE: 1996-06-14
; PRIOR APPLICATION NUMBER: 60/000,254
; PRIOR FILING DATE: 1995-06-15
; NUMBER OF SEQ ID NOS: 169
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 131
; LENGTH: 943
; TYPE: PRT
; ORGANISM: Mycobacterium tuberculosis
US-09-997-182-131

Query Match      11.4%; Score 86.5; DB 10; Length 943;
Best Local Similarity 24.8%; Pred. No. 3.1;
Matches 39; Conservative 22; Mismatches 73; Indels 23; Gaps 6;

QY 7 AAFPAAIIVSG-----SALAGVDLVTRVVTTHMAHAGSPDSTLS-IYQYGSANAA 56
   |||||
Db 96 ALFVSLVSNLGNAPALAAATEAAVEQWMAQDVAAAMFGHAGASAASALTFFGQALPT 155
   |||||
QY 57 L-----ALQSDARKSETTITQSGYNGADYGGADNSTIETQNGFRNNAIDQWNAKNSD 112
   |||||
Db 156 VAGGALVSAQAQVTRVFRNLGL-ANVGEGN-----VGNVGNFNLGSANIGN 207
   |||||
QY 113 ITVQYGGNNAALVNOTASDSSVMVRQVGFNNATAN 149
   |||||
Db 208 IGSNGISSNIGFCN-VGPCLTAALNNIGFGNTGSNN 243

RESULT 13
US-09-997-181-131
; Sequence 131, Application US/09997181
; Publication No. US20030049269A1
; GENERAL INFORMATION:
; APPLICANT: Nano, Francis
; TITLE OF INVENTION: Mycobacterium Tuberculosis DNA Sequences Encoding
; TITLE OF INVENTION: immunostimulatory Peptides
; FILE REFERENCE: 61257
; CURRENT APPLICATION NUMBER: US/09/997,181
; CURRENT FILING DATE: 2001-11-28
; PRIOR APPLICATION NUMBER: 09/447,135
; PRIOR FILING DATE: 2000-01-03
; PRIOR APPLICATION NUMBER: 08/990,823
; PRIOR FILING DATE: 1997-12-15
; PRIOR APPLICATION NUMBER: US 96/10375
; PRIOR FILING DATE: 1996-06-14
; PRIOR APPLICATION NUMBER: 60/000,254
; PRIOR FILING DATE: 1995-06-15
; NUMBER OF SEQ ID NOS: 169
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 131
; LENGTH: 943
; TYPE: PRT
; ORGANISM: Mycobacterium tuberculosis
US-09-997-181-131

Query Match      11.4%; Score 86.5; DB 10; Length 943;
Best Local Similarity 24.8%; Pred. No. 3.1;
Matches 39; Conservative 22; Mismatches 73; Indels 23; Gaps 6;
```

Qy	7	A	P	A	P	A	I	V	S	G	-----	S	A	L	A	G	Y	D	O	L	T	R	V	T	H	E	M	A	S	G	P	D	S	T	L	S	-	I	Y	O	G	S	A	N	A	A	56						
Db	96	A	L	F	U	S	W	S	N	L	G	Q	N	A	P	A	I	E	A	E	O	M	Q	A	D	V	A	A	F	G	H	A	G	A	S	A	V	S	A	L	T	F	F	G	A	L	T	155					
Qy	57	L	-----	A	L	O	S	D	A	R	K	S	E	T	T	I	Q	S	G	Y	G	N	A	D	V	G	O	G	A	D	N	S	T	I	E	L	T	Q	N	F	R	N	A	I	D	O	M	N	A	K	S	D	112
Db	156	V	A	G	G	A	L	V	A	A	A	A	A	Q	V	T	T	R	N	L	G	L	-	A	N	V	G	E	N	-----	V	G	N	G	V	F	N	L	G	S	A	N	I	G	N	G	207						
Qy	113	I	T	V	Q	Y	G	C	G	N	N	A	L	V	N	O	T	A	S	D	S	V	M	V	R	V	G	F	G	N	N	A	T	A	N	149																	
Db	208	I	G	S	G	N	I	G	S	N	I	G	F	G	N	-	V	G	P	G	L	T	A	A	L	N	I	G	F	G	T	G	S	N	243																		

RESULT 14  
US-10-282-122A-64369  
Sequence 64369, Application US/10282122A  
Publication No. US20040029129A1  
GENERAL INFORMATION:  
APPLICANT: Wang, Liangsuo  
APPLICANT: Zamudio, Carlos  
APPLICANT: Malone, Cheryl  
APPLICANT: Haselbeck, Robert  
APPLICANT: Ohlsen, Karl  
APPLICANT: Zyskind, Judith  
APPLICANT: Wall, Daniel  
APPLICANT: Trawick, John  
APPLICANT: Carr, Grant  
APPLICANT: Yamamoto, Robert  
APPLICANT: Forsyth, R.  
APPLICANT: Xu, H.  
TITLE OF INVENTION: Identification of Essential Genes in Microorganisms  
FILE REFERENCE: ELITRA.034A  
CURRENT APPLICATION NUMBER: US/10/282,122A  
CURRENT FILING DATE: 2003-02-20  
PRIOR APPLICATION NUMBER: 60/191,078  
PRIOR FILING DATE: 2000-03-21  
PRIOR APPLICATION NUMBER: 60/206,848  
PRIOR FILING DATE: 2000-05-23  
PRIOR APPLICATION NUMBER: 60/207,727  
PRIOR FILING DATE: 2000-05-26  
PRIOR APPLICATION NUMBER: 60/230,335  
PRIOR FILING DATE: 2000-09-06  
PRIOR APPLICATION NUMBER: 60/230,347  
PRIOR FILING DATE: 2000-09-09  
PRIOR APPLICATION NUMBER: 60/242,578  
PRIOR FILING DATE: 2000-10-23  
PRIOR APPLICATION NUMBER: 60/253,625  
PRIOR FILING DATE: 2000-11-27  
PRIOR APPLICATION NUMBER: 60/257,931  
PRIOR FILING DATE: 2000-12-22  
PRIOR APPLICATION NUMBER: 60/267,636  
PRIOR FILING DATE: 2001-02-09  
PRIOR APPLICATION NUMBER: 60/269,308  
PRIOR FILING DATE: 2001-02-16  
Remaining prior Application data removed - See File Wrapper or PALM.  
NUMBER OF SEQ ID NOS: 78614  
SOFTWARE: PatentIn version 3.1  
SEQ ID NO 64369  
LENGTH: 3300  
TYPE: PRT  
ORGANISM: Mycobacterium tuberculosis  
US-10-282-122A-64369

Db 175 VAGGALVSRAAAQVTTTFVFNGL-ANVGEGN-----VGNVGNVFNLSANIGCN 228

Qy 113 ITVGYGNNAAALVNQTASDSSVMVRGVGFNNATAN 149

Db 227 IGSGNIGSSNIGFGN-VGPGHATAALNNIGFTGNSNN 262

```

RESULT 15
US-10-383-930-38
; Sequence 38, Application US/10383930
; Publication NO. US20040127400A1
; GENERAL INFORMATION:
; APPLICANT: Smith, Daniel J
; APPLICANT: Taubman, Martin A
; TITLE OF INVENTION: Immunogenicity of Glucan Binding Protein
; FILE REFERENCE: 25669-018
; CURRENT APPLICATION NUMBER: US/10/383,930
; CURRENT FILING DATE: 2003-03-07
; PRIOR APPLICATION NUMBER: 60/402,483
; PRIOR FILING DATE: 2002-08-08
; PRIOR APPLICATION NUMBER: 60/363,209
; PRIOR FILING DATE: 2002-03-07
; NUMBER OF SEQ ID NOS: 41
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 38
; LENGTH: 1554
; TYPE: PRT
; ORGANISM: Streptococcus sobrinus
US-10-383-930-38

```

Search completed: August 2, 2004, 15:36:10  
Job time : 37.8 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: August 2, 2004, 14:48:33 ; Search time 167.9 Seconds  
(without alignments)  
877.809 Million cell updates/sec

Title: US-09-543-407-16

Perfect score: 757

Sequence: 1 MKLLKVAFAAIVVSGSALA.....DSSVMVRQVGFNNATANQY 151

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 6019581 seqs, 976053577 residues

Total number of hits satisfying chosen parameters: 6019581

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

Pending Patents AA\_Main:\*

- 1: /cgn2\_6/prodata/2/paa/pctus COMB.pep.\*
- 2: /cgn2\_6/prodata/2/paa/US06 COMB.pep.\*
- 3: /cgn2\_6/prodata/2/paa/US07 COMB.pep.\*
- 4: /cgn2\_6/prodata/2/paa/US08 COMB.pep.\*
- 5: /cgn2\_6/prodata/2/paa/US081 COMB.pep.\*
- 6: /cgn2\_6/prodata/2/paa/US082 COMB.pep.\*
- 7: /cgn2\_6/prodata/2/paa/US083 COMB.pep.\*
- 8: /cgn2\_6/prodata/2/paa/US084 COMB.pep.\*
- 9: /cgn2\_6/prodata/2/paa/US085 COMB.pep.\*
- 10: /cgn2\_6/prodata/2/paa/US086 COMB.pep.\*
- 11: /cgn2\_6/prodata/2/paa/US087 COMB.pep.\*
- 12: /cgn2\_6/prodata/2/paa/US088 COMB.pep.\*
- 13: /cgn2\_6/prodata/2/paa/US089 COMB.pep.\*
- 14: /cgn2\_6/prodata/2/paa/US090 COMB.pep.\*
- 15: /cgn2\_6/prodata/2/paa/US091 COMB.pep.\*
- 16: /cgn2\_6/prodata/2/paa/US092 COMB.pep.\*
- 17: /cgn2\_6/prodata/2/paa/US093 COMB.pep.\*
- 18: /cgn2\_6/prodata/2/paa/US094 COMB.pep.\*
- 19: /cgn2\_6/prodata/2/paa/US095 COMB.pep.\*
- 20: /cgn2\_6/prodata/2/paa/US096 COMB.pep.\*
- 21: /cgn2\_6/prodata/2/paa/US097A COMB.pep.\*
- 22: /cgn2\_6/prodata/2/paa/US097B COMB.pep.\*
- 23: /cgn2\_6/prodata/2/paa/US098 COMB.pep.\*
- 24: /cgn2\_6/prodata/2/paa/US099A COMB.pep.\*
- 25: /cgn2\_6/prodata/2/paa/US099B COMB.pep.\*
- 26: /cgn2\_6/prodata/2/paa/US100 COMB.pep.\*
- 27: /cgn2\_6/prodata/2/paa/US101 COMB.pep.\*
- 28: /cgn2\_6/prodata/2/paa/US102 COMB.pep.\*
- 29: /cgn2\_6/prodata/2/paa/US103 COMB.pep.\*
- 30: /cgn2\_6/prodata/2/paa/US104 COMB.pep.\*
- 31: /cgn2\_6/prodata/2/paa/US106 COMB.pep.\*
- 32: /cgn2\_6/prodata/2/paa/US107 COMB.pep.\*
- 33: /cgn2\_6/prodata/2/paa/US60 COMB.pep.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match Length	ID	Description
------------	-------	--------------------	----	-------------

1	757	100.0	151	19	US-09-543-407-16	Sequence 16, Appl
2	658	86.9	151	19	US-09-543-407-5	Sequence 5, Appl
3	653	86.3	151	6	US-08-233-642A-57	Sequence 57, Appl
4	618	81.6	151	19	US-09-543-407-18	Sequence 18, Appl
5	580	76.6	151	19	US-09-543-407-12	Sequence 12, Appl
6	579	76.5	131	19	US-09-543-407-31	Sequence 31, Appl
7	578	76.4	151	19	US-09-543-407-14	Sequence 14, Appl
8	577	76.2	151	19	US-09-543-407-26	Sequence 26, Appl
9	575	76.0	151	19	US-09-543-407-24	Sequence 24, Appl
10	574	75.8	151	19	US-09-543-407-20	Sequence 20, Appl
11	568	75.0	151	19	US-09-543-407-28	Sequence 28, Appl
12	567	74.9	151	19	US-09-543-407-22	Sequence 22, Appl
13	566	74.8	151	19	US-09-543-407-30	Sequence 30, Appl
14	554	73.2	109	19	US-09-543-407-34	Sequence 34, Appl
15	487	64.3	120	6	US-08-233-642A-55	Sequence 55, Appl
16	484	63.9	151	19	US-09-543-407-7	Sequence 7, Appl
17	481	63.5	151	13	US-08-978-878-4	Sequence 4, Appl
18	479	63.3	151	21	US-09-741-873B-4	Sequence 4, Appl
19	479	63.3	151	33	US-60-352-946-2	Sequence 2, Appl
20	479	63.3	151	33	US-60-444-371-2	Sequence 2, Appl
21	468	61.8	158	16	US-09-252-691-5834	Sequence 5834, Ap
22	468	61.8	158	16	US-09-252-691C-5834	Sequence 5834, Ap
23	468	61.8	158	20	US-10-417-886-5834	Sequence 5834, Ap
24	414	54.7	131	13	US-08-978-878-2	Sequence 2, Appl
25	414	54.7	131	21	US-09-741-873B-2	Sequence 2, Appl
26	401	53.0	109	19	US-09-543-407-35	Sequence 35, Appl
27	347	45.8	68	19	US-09-543-407-37	Sequence 37, Appl
28	250.5	33.1	70	19	US-09-543-407-32	Sequence 32, Appl
29	237	31.3	48	19	US-09-543-407-39	Sequence 39, Appl
30	122.5	16.2	186	16	US-09-252-691-5833	Sequence 5833, Ap
31	122.5	16.2	186	16	US-09-252-691C-5833	Sequence 5833, Ap
32	122.5	16.2	186	30	US-10-417-886-5833	Sequence 5833, Ap
33	118.5	15.7	151	19	US-09-543-407-8	Sequence 8, Appl
34	113.5	15.0	445	29	US-10-369-493-20638	Sequence 20638, A
35	113.5	15.0	445	33	US-60-360-039-20638	Sequence 20638, A
36	111.5	14.7	151	19	US-09-543-407-6	Sequence 6, Appl
37	101.5	13.4	590	3	US-07-965-673A-10	Sequence 10, Appl
38	101.5	13.4	590	22	US-09-791-537-29229	Sequence 29229, A
39	101.5	13.4	590	22	US-09-791-537-111372	Sequence 111372, A
40	101	13.3	599	22	US-09-791-537-29228	Sequence 29228, A
41	101	13.3	599	22	US-09-791-537-111371	Sequence 111371, A
42	99.5	13.1	582	22	US-09-791-537-94843	Sequence 94843, A
43	99.5	13.1	646	22	US-09-791-537-30765	Sequence 30765, A
44	99.5	13.1	646	22	US-09-791-537-32862	Sequence 32862, A
45	98	12.9	145	21	US-09-739-449-8854	Sequence 8854, Ap

#### ALIGNMENTS

RESULT 1  
US-09-543-407-16  
; Sequence 16, Application US/09543407  
; GENERAL INFORMATION:  
; APPLICANT: White, Aaron P.  
; APPLICANT: Doran, James L.  
; APPLICANT: Collinson, S. Karen  
; APPLICANT: Kay, William W.  
; TITLE OF INVENTION: BACTERIAL FMBRIAL SYSTEM FOR  
; TITLE OF INVENTION: PRESENTATION OF HETEROLOGOUS PEPTIDE SEQUENCES  
; FILE REFERENCE: 920043.406  
; CURRENT APPLICATION NUMBER: US/09/543.407  
; CURRENT FILING DATE: 2000-04-05  
; NUMBER OF SEQ ID NOS: 59  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 16  
; LENGTH: 151  
; TYPE: PRT  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: Recombinant Salmonella enteritidis 3b afga  
; OTHER INFORMATION: Sequence containing the replacement fragment  
; OTHER INFORMATION: encoding PT3 from GP63 of Leishmania major.

US-09-543-407-16

Query Match 100.0%; Score 757; DB 19; Length 151;  
 Best Local Similarity 100.0%; Pred. No. 3.4e-76;  
 Matches 151; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MKLLKVAAPAAIIVVSGSALAGVVDQLVTRVTHMAHAGSPDSTLSIYQYGSANAALALQ 60

DB 1 MKLLKVAAPAAIIVVSGSALAGVVDQLVTRVTHMAHAGSPDSTLSIYQYGSANAALALQ 60

QY 61 SDARKSETTITQSGYNGADVGAGDNSTIELTQNGFRNNATIDQNAKNSDITVGYGG 120

DB 61 SDARKSETTITQSGYNGADVGAGDNSTIELTQNGFRNNATIDQNAKNSDITVGYGG 120

QY 121 NNAALVNOTASDSSVMVRQVGFNNATANOY 151

DB 121 NNAALVNOTASDSSVMVRQVGFNNATANOY 151

## RESULT 2

US-09-543-407-5

; Sequence 5, Application US/09543407

; GENERAL INFORMATION:

; APPLICANT: White, Aaron P.

; APPLICANT: Doran, James L.

; APPLICANT: Collinson, S. Karen

; APPLICANT: Kay, William W.

; TITLE OF INVENTION: BACTERIAL FIMBRIAL SYSTEM FOR

; PRESENTATION OF HETEROLOGOUS PEPTIDE SEQUENCES

; FILE REFERENCE: 920043.406

; CURRENT APPLICATION NUMBER: US/09/543,407

; CURRENT FILING DATE: 2000-04-05

; NUMBER OF SEQ ID NOS: 59

; SOFTWARE: FastSeq for Windows Version 4.0

; SEQ ID NO 5

; LENGTH: 151

; TYPE: PRT

; ORGANISM: Salmonella enteritidis

US-09-543-407-5

Query Match 86.9%; Score 658; DB 19; Length 151;  
 Best Local Similarity 90.1%; Pred. No. 4.5e-65;  
 Matches 136; Conservative 3; Mismatches 12; Indels 0; Gaps 0;

QY 1 MKLLKVAAPAAIIVVSGSALAGVVDQLVTRVTHMAHAGSPDSTLSIYQYGSANAALALQ 60

DB 1 MKLLKVAAPAAIIVVSGSALAGVVDQLVTRVTHMAHAGSPDSTLSIYQYGSANAALALQ 60

QY 61 SDARKSETTITQSGYNGADVGAGDNSTIELTQNGFRNNATIDQNAKNSDITVGYGG 120

DB 61 SDARKSETTITQSGYNGADVGAGDNSTIELTQNGFRNNATIDQNAKNSDITVGYGG 120

QY 121 NNAALVNOTASDSSVMVRQVGFNNATANOY 151

DB 121 NNAALVNOTASDSSVMVRQVGFNNATANOY 151

## RESULT 3

US-08-233-642A-57

; Sequence 57, Application US/08233642A

; GENERAL INFORMATION:

; APPLICANT: Kay, William W.

; APPLICANT: Collinson, S. Karen

; APPLICANT: Clouthier, Sharon C.

; APPLICANT: Doran, James L.

; TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR SALMONELLA-

; BASED VACCINES

; NUMBER OF SEQUENCES: 58

; NUMBER OF SEQUENCES: 58

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Seed and Berry

; STREET: 6300 Columbia Center, 701 Fifth Avenue

; CITY: Seattle

; STATE: Washington  
 ; COUNTRY: U.S.A.  
 ; ZIP: 98104-7092  
 ; COMPUTER READABLE FORM:  
 ; MEDIUM TYPE: Floppy disk  
 ; COMPUTER: IBM PC compatible  
 ; OPERATING SYSTEM: PC-DOS/MS-DOS  
 ; SOFTWARE: PatentIn Release #1.0, Version #1.25  
 ; CURRENT APPLICATION DATA: US/08/233,642A  
 ; APPLICATION NUMBER: US/08/233,642A  
 ; FILING DATE: 26-APR-1994  
 ; CLASSIFICATION: 424  
 ; ATTORNEY/AGENT INFORMATION:  
 ; NAME: King, Joshua  
 ; REGISTRATION NUMBER: 35,570  
 ; REFERENCE/DOCKET NUMBER: 920043.403C3  
 ; TELECOMMUNICATION INFORMATION:  
 ; TELEPHONE: (206) 622-4900  
 ; TELEFAX: (206) 682-6031  
 ; TELEX: 3723836 SEEDANBERRY  
 ; INFORMATION FOR SEQ ID NO: 57:  
 ; SEQUENCE CHARACTERISTICS:  
 ; LENGTH: 151 amino acids  
 ; TYPE: amino acid  
 ; TOPOLOGY: linear  
 ; MOLECULE TYPE: protein  
 ; US-08-233-642A-57

Query Match 86.3%; Score 653; DB 6; Length 151;

Best Local Similarity 89.4%; Pred. No. 1.6e-64;

Matches 135; Conservative 3; Mismatches 13; Indels 0; Gaps 0;

QY 1 MKLLKVAAPAAIIVVSGSALAGVVDQLVTRVTHMAHAGSPDSTLSIYQYGSANAALALQ 60

DB 1 MKLLKVAAPAAIIVVSGSALAGVVDQLVTRVTHMAHAGSPDSTLSIYQYGSANAALALQ 60

QY 61 SDARKSETTITQSGYNGADVGAGDNSTIELTQNGFRNNATIDQNAKNSDITVGYGG 120

DB 61 SDARKSETTITQSGYNGADVGAGDNSTIELTQNGFRNNATIDQNAKNSDITVGYGG 120

QY 121 NNAALVNOTASDSSVMVRQVGFNNATANOY 151

DB 121 NNAALVNOTASDSSVMVRQVGFNNATANOY 151

## RESULT 4

US-09-543-407-18

; Sequence 18, Application US/09543407

; GENERAL INFORMATION:

; APPLICANT: White, Aaron P.

; APPLICANT: Doran, James L.

; APPLICANT: Collinson, S. Karen

; APPLICANT: Kay, William W.

; TITLE OF INVENTION: BACTERIAL FIMBRIAL SYSTEM FOR

; PRESENTATION OF HETEROLOGOUS PEPTIDE SEQUENCES

; FILE REFERENCE: 920043.406

; CURRENT APPLICATION NUMBER: US/09/543,407

; CURRENT FILING DATE: 2000-04-05

; NUMBER OF SEQ ID NOS: 59

; SOFTWARE: FastSeq for Windows Version 4.0

; SEQ ID NO 18

; LENGTH: 151

; TYPE: PRT

; ORGANISM: Artificial Sequence

; FEATURE:

; OTHER INFORMATION: Recombinant Salmonella enteritidis 3b afga

; OTHER INFORMATION: sequence containing the replacement fragment

; OTHER INFORMATION: encoding P73 from GPe3 of Leishmania major.

US-09-543-407-18

Query Match 81.6%; Score 618; DB 19; Length 151;

Best Local Similarity 76.6%; Pred. No. 1.4e-60;

Matches 131; Conservative 0; Mismatches 0; Indels 40; Gaps 2;

```
QY 1 MKLLKVAFAAIVVSGSALAGV-----YDQLVTRVVTHEMAHSG 40
DQ 1 MKLLKVAFAAIVVSGSALAGVVPQWGGGNGHNGSGPDDYDQLVTRVVTHEMAH--- 57
QY 41 PDSTLSIYQGSANAALQSDARKSETTITQSGYNGADVGQADNSTIETQNGFRNN 100
DQ 58 -----ALQSDARKSETTITQSGYNGADVGQADNSTIETQNGFRNN 100
QY 101 ATIDQWNAKNSDITVQGYGNNALVNQTSASSVMVRQVGFNNATANY 151
DQ 101 ATIDQWNAKNSDITVQGYGNNALVNQTSASSVMVRQVGFNNATANY 151

RESULT 5
US-09-543-407-12
; Sequence 12, Application US/09543407
; GENERAL INFORMATION:
; APPLICANT: White, Aaron P.
; APPLICANT: Doran, James L.
; APPLICANT: Collinson, S. Karen
; APPLICANT: Kay, William W.
; TITLE OF INVENTION: BACTERIAL FIMBRIAL SYSTEM FOR
; TITLE OF INVENTION: PRESENTATION OF HETEROLOGOUS PEPTIDE SEQUENCES
; FILE REFERENCE: 920043.406
; CURRENT APPLICATION NUMBER: US/09/543,407
; CURRENT FILING DATE: 2000-04-05
; NUMBER OF SEQ ID NOS: 59
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 12
; LENGTH: 151
; ORGANISM: Artificial Sequence
; TYPE: PRP
; FEATURE:
; OTHER INFORMATION: Recombinant Salmonella enteritidis 3b afgA
; OTHER INFORMATION: sequence containing the replacement fragment
; OTHER INFORMATION: encoding PT3 from GP63 of Leishmania major.
US-09-543-407-12

Query Match 76.6%; Score 580; DB 19; Length 151;
Best Local Similarity 80.1%; Pred. No. 2.6e-56;
Matches 121; Conservative 7; Mismatches 23; Indels 0; Gaps 0;

QY 1 MKLLKVAFAAIVVSGSALAGVYDQLVTRVVTHEMAHSGPDSSTLSIYQGSANAALQ 60
DQ 1 MKLLKVAFAAIVVSGSALAGVVPQWGGGNGHNGSGPDSSTLSIYQGSANAALQ 60
QY 61 SDARKSETTITQSGYNGADVGQADNSTIETQNGFRNNATIDQWNAKNSDITVQYGG 120
DQ 61 SDARKSETTITQSGYNGADVGQADNSTIETQNGFRNNATIDQWNAKNSDITVQYGG 120
QY 121 NNAALVNQTSASSVMVRQVGFNNATANY 151
DQ 121 NNAALVNQTSASSVMVRQVGFNNATANY 151

RESULT 6
US-09-543-407-31
; Sequence 31, Application US/09543407
; GENERAL INFORMATION:
; APPLICANT: White, Aaron P.
; APPLICANT: Doran, James L.
; APPLICANT: Collinson, S. Karen
; APPLICANT: Kay, William W.
; TITLE OF INVENTION: BACTERIAL FIMBRIAL SYSTEM FOR
; TITLE OF INVENTION: PRESENTATION OF HETEROLOGOUS PEPTIDE SEQUENCES
; FILE REFERENCE: 920043.406
; CURRENT APPLICATION NUMBER: US/09/543,407
; CURRENT FILING DATE: 2000-04-05
; NUMBER OF SEQ ID NOS: 59
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 31
; LENGTH: 131
```

```
; TYPE: PRP
; ORGANISM: Salmonella enteritidis
US-09-543-407-31

Query Match 76.5%; Score 579; DB 19; Length 131;
Best Local Similarity 98.3%; Pred. No. 2.8e-56;
Matches 113; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 37 HASGPDSTLSIYQGSANAALQSDARKSETTITQSGYNGADVGQADNSTIETQNG 96
DQ 17 NSSGPDSTLSIYQGSANAALQSDARKSETTITQSGYNGADVGQADNSTIETQNG 76
QY 97 FRNNATIDQWNAKNSDITVQGYGNNALVNQTSASSVMVRQVGFNNATANY 151
DQ 77 FRNNATIDQWNAKNSDITVQGYGNNALVNQTSASSVMVRQVGFNNATANY 131

RESULT 7
US-09-543-407-14
; Sequence 14, Application US/09543407
; GENERAL INFORMATION:
; APPLICANT: White, Aaron P.
; APPLICANT: Doran, James L.
; APPLICANT: Collinson, S. Karen
; APPLICANT: Kay, William W.
; TITLE OF INVENTION: BACTERIAL FIMBRIAL SYSTEM FOR
; TITLE OF INVENTION: PRESENTATION OF HETEROLOGOUS PEPTIDE SEQUENCES
; FILE REFERENCE: 920043.406
; CURRENT APPLICATION NUMBER: US/09/543,407
; CURRENT FILING DATE: 2000-04-05
; NUMBER OF SEQ ID NOS: 59
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 14
; LENGTH: 151
; ORGANISM: Artificial Sequence
; TYPE: PRP
; FEATURE:
; OTHER INFORMATION: Recombinant Salmonella enteritidis 3b afgA
; OTHER INFORMATION: sequence containing the replacement fragment
; OTHER INFORMATION: encoding PT3 from GP63 of Leishmania major.
US-09-543-407-14

Query Match 76.4%; Score 578; DB 19; Length 151;
Best Local Similarity 80.8%; Pred. No. 4.4e-56;
Matches 122; Conservative 6; Mismatches 23; Indels 0; Gaps 0;

QY 1 MKLLKVAFAAIVVSGSALAGVYDQLVTRVVTHEMAHSGPDSSTLSIYQGSANAALQ 60
DQ 1 MKLLKVAFAAIVVSGSALAGVVPQWGGGNGHNGSGPDSSTLSIYQGSANAALQ 60
QY 61 SDARKSETTITQSGYNGADVGQADNSTIETQNGFRNNATIDQWNAKNSDITVQYGG 120
DQ 61 SDARKSETTITQSGYNGADVGQADNSTIETQNGFRNNATIDQWNAKNSDITVQYGG 120
QY 121 NNAALVNQTSASSVMVRQVGFNNATANY 151
DQ 121 LVTRVVTHEMAHSGVVRQVGFNNATANY 151

RESULT 8
US-09-543-407-26
; Sequence 26, Application US/09543407
; GENERAL INFORMATION:
; APPLICANT: White, Aaron P.
; APPLICANT: Doran, James L.
; APPLICANT: Collinson, S. Karen
; APPLICANT: Kay, William W.
; TITLE OF INVENTION: BACTERIAL FIMBRIAL SYSTEM FOR
; TITLE OF INVENTION: PRESENTATION OF HETEROLOGOUS PEPTIDE SEQUENCES
; FILE REFERENCE: 920043.406
; CURRENT APPLICATION NUMBER: US/09/543,407
; CURRENT FILING DATE: 2000-04-05
; NUMBER OF SEQ ID NOS: 59
```

```
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 26
; LENGTH: 151
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Recombinant Salmonella enteritidis 3b afga
; OTHER INFORMATION: sequence containing the replacement fragment
; OTHER INFORMATION: encoding PT3 from GP63 of Leishmania major.
US-09-543-407-26

Query Match      76.2%; Score 577; DB 19; Length 151;
Best Local Similarity 80.8%; Pred. No. 5.7e-56;
Matches 122; Conservative 5; Mismatches 24; Indels 0; Gaps 0;

QY 1 MKLLKVAAPAAIIVVSGSALAGVVDQLVTRVVTHEMAHAGSPDSTLSIYQYGSANAALALQ 60
   |||||
Db 1 MKLLKVAAPAAIIVVSGSALAGVVPQWGGGNGHNGSGSPDSTLSIYQYGSANAALYDQ 60
   |||||
QY 61 SDARKSETTITQSGYNGADVGQADNSTIELTQNGFRNNATIDQWNAKNSDITVGYGG 120
   |||||
Db 61 LVTRVVTHEMAHAGYNGADVGQADNSTIELTQNGFRNNATIDQWNAKNSDITVGYGG 120
   |||||
QY 121 NNAALVNQTASDSSVMVRQVGFNNATANQY 151
   |||||
Db 121 NNAALVNQTASDSSVMVRQVGFNNATANQY 151
   |||||
```

```
RESULT 9
US-09-543-407-24
; Sequence 24, Application US/09543407
; GENERAL INFORMATION:
; APPLICANT: White, Aaron P.
; APPLICANT: Doran, James L.
; APPLICANT: Collinson, S. Karen
; APPLICANT: Kay, William W.
; TITLE OF INVENTION: BACTERIAL FIMBRIAL SYSTEM FOR
; FILE REFERENCE: 920043.406
; CURRENT APPLICATION NUMBER: US/09/543,407
; CURRENT FILING DATE: 2000-04-05
; NUMBER OF SEQ ID NOS: 59
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 24
; LENGTH: 151
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Recombinant Salmonella enteritidis 3b afga
; OTHER INFORMATION: sequence containing the replacement fragment
; OTHER INFORMATION: encoding PT3 from GP63 of Leishmania major.
US-09-543-407-24

Query Match      76.0%; Score 575; DB 19; Length 151;
Best Local Similarity 81.5%; Pred. No. 9.5e-56;
Matches 123; Conservative 5; Mismatches 23; Indels 0; Gaps 0;

QY 1 MKLLKVAAPAAIIVVSGSALAGVVDQLVTRVVTHEMAHAGSPDSTLSIYQYGSANAALALQ 60
   |||||
Db 1 MKLLKVAAPAAIIVVSGSALAGVVPQWGGGNGHNGSGSPDSTLSIYQYGSANAALALQ 60
   |||||
QY 61 SDARKSETTITQSGYNGADVGQADNSTIELTQNGFRNNATIDQWNAKNSDITVGYGG 120
   |||||
Db 61 SDARKSETTITQSGYNGADVGQADNSTIELTQNGFRNNATIDQWNAKNSDITVGYGG 120
   |||||
QY 121 NNAALVNQTASDSSVMVRQVGFNNATANQY 151
   |||||
Db 121 HEMAHAQNTASDSSVMVRQVGFNNATANQY 151
   |||||
```

```
RESULT 10
US-09-543-407-20
; Sequence 20, Application US/09543407
```

```
; GENERAL INFORMATION:
; APPLICANT: White, Aaron P.
; APPLICANT: Doran, James L.
; APPLICANT: Collinson, S. Karen
; APPLICANT: Kay, William W.
; TITLE OF INVENTION: BACTERIAL FIMBRIAL SYSTEM FOR
; FILE REFERENCE: 920043.406
; CURRENT APPLICATION NUMBER: US/09/543,407
; CURRENT FILING DATE: 2000-04-05
; NUMBER OF SEQ ID NOS: 59
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 20
; LENGTH: 151
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Recombinant Salmonella enteritidis 3b afga
; OTHER INFORMATION: sequence containing the replacement fragment
; OTHER INFORMATION: encoding PT3 from GP63 of Leishmania major.
US-09-543-407-20

Query Match      75.8%; Score 574; DB 19; Length 151;
Best Local Similarity 80.8%; Pred. No. 1.2e-55;
Matches 122; Conservative 6; Mismatches 23; Indels 0; Gaps 0;

QY 1 MKLLKVAAPAAIIVVSGSALAGVVDQLVTRVVTHEMAHAGSPDSTLSIYQYGSANAALALQ 60
   |||||
Db 1 MKLLKVAAPAAIIVVSGSALAGVVPQWGGGNGHNGSGSPDSTLSIYQYGSANAALALQ 60
   |||||
QY 61 SDARKSETTITQSGYNGADVGQADNSTIELTQNGFRNNATIDQWNAKNSDITVGYGG 120
   |||||
Db 61 SDARKYDQLVTRVVTHEMAHAGQADNSTIELTQNGFRNNATIDQWNAKNSDITVGYGG 120
   |||||
QY 121 NNAALVNQTASDSSVMVRQVGFNNATANQY 151
   |||||
Db 121 NNAALVNQTASDSSVMVRQVGFNNATANQY 151
   |||||
```

```
RESULT 11
US-09-543-407-28
; Sequence 28, Application US/09543407
; GENERAL INFORMATION:
; APPLICANT: White, Aaron P.
; APPLICANT: Doran, James L.
; APPLICANT: Collinson, S. Karen
; APPLICANT: Kay, William W.
; TITLE OF INVENTION: BACTERIAL FIMBRIAL SYSTEM FOR
; FILE REFERENCE: 920043.406
; CURRENT APPLICATION NUMBER: US/09/543,407
; CURRENT FILING DATE: 2000-04-05
; NUMBER OF SEQ ID NOS: 59
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 28
; LENGTH: 151
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Recombinant Salmonella enteritidis 3b afga
; OTHER INFORMATION: sequence containing the replacement fragment
; OTHER INFORMATION: encoding PT3 from GP63 of Leishmania major.
US-09-543-407-28

Query Match      75.0%; Score 568; DB 19; Length 151;
Best Local Similarity 80.1%; Pred. No. 5.8e-55;
Matches 121; Conservative 4; Mismatches 26; Indels 0; Gaps 0;

QY 1 MKLLKVAAPAAIIVVSGSALAGVVDQLVTRVVTHEMAHAGSPDSTLSIYQYGSANAALALQ 60
   |||||
Db 1 MKLLKVAAPAAIIVVSGSALAGVVPQWGGGNGHNGSGSPDSTLSIYQYGSANAALALQ 60
   |||||
QY 61 SDARKSETTITQSGYNGADVGQADNSTIELTQNGFRNNATIDQWNAKNSDITVGYGG 120
   |||||
```



```
Db 61 SDARKSETTITQSGYNGADYDQLVTRVVTHEMAHAFENATIDQWNAKNSDITVGOYGG 120
QY 121 NNAALVNQTSASSVWVRQVGFNNATANOY 151
Db 121 NNAALVNQTSASSVWVRQVGFNNATANOY 151

RESULT 12
US-09-543-407-22
; Sequence 22, Application US/09543407
; GENERAL INFORMATION:
; APPLICANT: White, Aaron P.
; APPLICANT: Doran, James L.
; APPLICANT: Collinson, S. Karen
; APPLICANT: Kay, William W.
; TITLE OF INVENTION: BACTERIAL FIMBRIAL SYSTEM FOR
; TITLE OF INVENTION: PRESENTATION OF HETEROLOGOUS PEPTIDE SEQUENCES
; FILE REFERENCE: 920043.406
; CURRENT APPLICATION NUMBER: US/09/543,407
; CURRENT FILING DATE: 2000-04-05
; NUMBER OF SEQ ID NOS: 59
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 22
; LENGTH: 151
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Recombinant Salmonella enteritidis 3b afga
; OTHER INFORMATION: sequence containing the replacement fragment
; OTHER INFORMATION: encoding PT3 from GP63 of Leishmania major.
US-09-543-407-22
Query Match 74.8%; Score 567; DB 19; Length 151;
Best Local Similarity 80.8%; Pred. No. 7.5e-55;
Matches 122; Conservative 5; Mismatches 24; Indels 0; Gaps 0;

QY 1 MKLLKVAAPFAAIIVVSGSALAGVYDQLVTRVVTHEMAHAGSPDSTLSIYQYGSANAALALQ 60
Db 1 MKLLKVAAPFAAIIVVSGSALAGVYDQLVTRVVTHEMAHAGSPDSTLSIYQYGSANAALALQ 60

QY 61 SDARKSETTITQSGYNGADVGQGANSTIELTQNGFRNNATIDQWNAKNSDITVGOYGG 120
Db 61 SDARKSETTITQSGYNGADVGQGANSTIELTQNGFRNNATIDQWNAKNSDITVGOYGG 120

QY 121 NNAALVNQTSASSVWVRQVGFNNATANOY 151
Db 121 NNAALVNQTSASSVWVRQVGFNNATANOY 151

RESULT 13
US-09-543-407-30
; Sequence 30, Application US/09543407
; GENERAL INFORMATION:
; APPLICANT: White, Aaron P.
; APPLICANT: Doran, James L.
; APPLICANT: Collinson, S. Karen
; APPLICANT: Kay, William W.
; TITLE OF INVENTION: BACTERIAL FIMBRIAL SYSTEM FOR
; TITLE OF INVENTION: PRESENTATION OF HETEROLOGOUS PEPTIDE SEQUENCES
; FILE REFERENCE: 920043.406
; CURRENT APPLICATION NUMBER: US/09/543,407
; CURRENT FILING DATE: 2000-04-05
; NUMBER OF SEQ ID NOS: 59
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 30
; LENGTH: 151
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Recombinant Salmonella enteritidis 3b afga
; OTHER INFORMATION: sequence containing the replacement fragment
; OTHER INFORMATION: encoding PT3 from GP63 of Leishmania major.
US-09-543-407-30
Query Match 74.8%; Score 567; DB 19; Length 151;
Best Local Similarity 80.8%; Pred. No. 7.5e-55;
Matches 122; Conservative 5; Mismatches 24; Indels 0; Gaps 0;

QY 1 MKLLKVAAPFAAIIVVSGSALAGVYDQLVTRVVTHEMAHAGSPDSTLSIYQYGSANAALALQ 60
Db 1 MKLLKVAAPFAAIIVVSGSALAGVYDQLVTRVVTHEMAHAGSPDSTLSIYQYGSANAALALQ 60

QY 61 SDARKSETTITQSGYNGADVGQGANSTIELTQNGFRNNATIDQWNAKNSDITVGOYGG 120
Db 61 SDARKSETTITQSGYNGADVGQGANSTIELTQNGFRNNATIDQWNAKNSDITVGOYGG 120

QY 121 NNAALVNQTSASSVWVRQVGFNNATANOY 151
Db 121 NNAALVNQTSASSVWVRQVGFNNATANOY 151

RESULT 14
US-09-543-407-34
; Sequence 34, Application US/09543407
; GENERAL INFORMATION:
; APPLICANT: White, Aaron P.
; APPLICANT: Doran, James L.
; APPLICANT: Collinson, S. Karen
; APPLICANT: Kay, William W.
; TITLE OF INVENTION: BACTERIAL FIMBRIAL SYSTEM FOR
; TITLE OF INVENTION: PRESENTATION OF HETEROLOGOUS PEPTIDE SEQUENCES
; FILE REFERENCE: 920043.406
; CURRENT APPLICATION NUMBER: US/09/543,407
; CURRENT FILING DATE: 2000-04-05
; NUMBER OF SEQ ID NOS: 59
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 34
; LENGTH: 109
; TYPE: PRT
; ORGANISM: Salmonella enteritidis
US-09-543-407-34
Query Match 73.2%; Score 554; DB 19; Length 109;
Best Local Similarity 100.0%; Pred. No. 1.4e-53;
Matches 109; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 43 STLSTIYQYGSANAALALQSDARKSETTITQSGYNGADVGQGANSTIELTQNGFRNNAT 102
Db 1 STLSTIYQYGSANAALALQSDARKSETTITQSGYNGADVGQGANSTIELTQNGFRNNAT 60

QY 103 IDQWNAKNSDITVGOYGGNNAALVNQTSASSVWVRQVGFNNATANOY 151
Db 61 IDQWNAKNSDITVGOYGGNNAALVNQTSASSVWVRQVGFNNATANOY 109

RESULT 15
US-08-233-642A-55
; Sequence 55, Application US/08233642A
; GENERAL INFORMATION:
; APPLICANT: Kay, William W.
; APPLICANT: Collinson, S. Karen
; APPLICANT: Clouthier, Sharon C.
; APPLICANT: Doran, James L.
; TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR SALMONELLA-
; TITLE OF INVENTION: BASED VACCINES
; NUMBER OF SEQUENCES: 58
; NUMBER OF SEQUENCES: -
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Seed and Berry
; STREET: 6300 Columbia Center, 701 Fifth Avenue
; CITY: Seattle
; STATE: Washington
; COUNTRY: U.S.A.
; ZIP: 98104-7092
; COMPUTER READABLE FORM:
```

```
US-09-543-407-30
Query Match 74.8%; Score 566; DB 19; Length 151;
Best Local Similarity 80.8%; Pred. No. 9.7e-55;
Matches 122; Conservative 5; Mismatches 24; Indels 0; Gaps 0;

QY 1 MKLLKVAAPFAAIIVVSGSALAGVYDQLVTRVVTHEMAHAGSPDSTLSIYQYGSANAALALQ 60
Db 1 MKLLKVAAPFAAIIVVSGSALAGVYDQLVTRVVTHEMAHAGSPDSTLSIYQYGSANAALALQ 60

QY 61 SDARKSETTITQSGYNGADVGQGANSTIELTQNGFRNNATIDQWNAKNSDITVGOYGG 120
Db 61 SDARKSETTITQSGYNGADVGQGANSTIELTQNGFRNNATIDQWNAKNSDITVGOYGG 120

QY 121 NNAALVNQTSASSVWVRQVGFNNATANOY 151
Db 121 NNAALVNQTSASSVWVRQVGFNNATANOY 151

RESULT 14
US-09-543-407-34
; Sequence 34, Application US/09543407
; GENERAL INFORMATION:
; APPLICANT: White, Aaron P.
; APPLICANT: Doran, James L.
; APPLICANT: Collinson, S. Karen
; APPLICANT: Kay, William W.
; TITLE OF INVENTION: BACTERIAL FIMBRIAL SYSTEM FOR
; TITLE OF INVENTION: PRESENTATION OF HETEROLOGOUS PEPTIDE SEQUENCES
; FILE REFERENCE: 920043.406
; CURRENT APPLICATION NUMBER: US/09/543,407
; CURRENT FILING DATE: 2000-04-05
; NUMBER OF SEQ ID NOS: 59
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 34
; LENGTH: 109
; TYPE: PRT
; ORGANISM: Salmonella enteritidis
US-09-543-407-34
Query Match 73.2%; Score 554; DB 19; Length 109;
Best Local Similarity 100.0%; Pred. No. 1.4e-53;
Matches 109; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 43 STLSTIYQYGSANAALALQSDARKSETTITQSGYNGADVGQGANSTIELTQNGFRNNAT 102
Db 1 STLSTIYQYGSANAALALQSDARKSETTITQSGYNGADVGQGANSTIELTQNGFRNNAT 60

QY 103 IDQWNAKNSDITVGOYGGNNAALVNQTSASSVWVRQVGFNNATANOY 151
Db 61 IDQWNAKNSDITVGOYGGNNAALVNQTSASSVWVRQVGFNNATANOY 109

RESULT 15
US-08-233-642A-55
; Sequence 55, Application US/08233642A
; GENERAL INFORMATION:
; APPLICANT: Kay, William W.
; APPLICANT: Collinson, S. Karen
; APPLICANT: Clouthier, Sharon C.
; APPLICANT: Doran, James L.
; TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR SALMONELLA-
; TITLE OF INVENTION: BASED VACCINES
; NUMBER OF SEQUENCES: 58
; NUMBER OF SEQUENCES: -
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Seed and Berry
; STREET: 6300 Columbia Center, 701 Fifth Avenue
; CITY: Seattle
; STATE: Washington
; COUNTRY: U.S.A.
; ZIP: 98104-7092
; COMPUTER READABLE FORM:
```

```

; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/233,642A
; FILING DATE: 26-APR-1994
; CLASSIFICATION: 424
; ATTORNEY/AGENT INFORMATION:
; NAME: King, Joshua
; REGISTRATION NUMBER: 35,570
; REFERENCE/DOCKET NUMBER: 920043.403C3
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (206) 622-4900
; TELEFAX: (206) 682-6031
; TELEX: 3723836 SEEDANBERRY
; INFORMATION FOR SEQ ID NO: 55:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 120 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-233-642A-55

Query Match      64.3%; Score 487; DB 6; Length 120;
Best Local Similarity 97.9%; Pred. No. 5.2e-46;
Matches 95; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 37 HASGPDSTLSIYQGSANAALALQSDARKSETTITQSGYNGADYVGQAGDNSTIELTQNG 96
   :|||||
Db 16 NSSGPDSTLSIYQGSANAALALQSDARKSETTITQSGYNGADYVGQAGDNSTIELTQNG 75
   :|||||

QY 97 FRNNATIDQNAKNSDITVQYGGNNAALVNOTASDS 133
   |||||
Db 76 FRNNATIDQNAKNSDITVQYGGNNAALVNOTASDS 112
   |||||
```

Search completed: August 2, 2004, 15:26:43  
Job time : 168.9 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: August 2, 2004, 14:49:38 ; Search time 17.8 seconds  
(without alignments)  
888.146 Million cell updates/sec

Title: US-09-543-407-16  
Perfect score: 757  
Sequence: 1 MLLKVAFAAIVVSGSALA.....DSSVMVROVQFGNNATANOY 151

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 601315 seqs, 104695340 residues

Total number of hits satisfying chosen parameters: 601315

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Pending Patents AA New:  
1: /cgn2\_6/ptodata/2/paa/PCT\_NEW\_COMB.pep.\*  
2: /cgn2\_6/ptodata/2/paa/US06\_NEW\_COMB.pep.\*  
3: /cgn2\_6/ptodata/2/paa/US07\_NEW\_COMB.pep.\*  
4: /cgn2\_6/ptodata/2/paa/US08\_NEW\_COMB.pep.\*  
5: /cgn2\_6/ptodata/2/paa/US09\_NEW\_COMB.pep.\*  
6: /cgn2\_6/ptodata/2/paa/US10\_NEW\_COMB.pep.\*  
7: /cgn2\_6/ptodata/2/paa/US60\_NEW\_COMB.pep.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	481	63.5	151	5	US-09-741-873C-4
2	414	54.7	131	5	US-09-741-873C-2
3	93	12.3	348	6	US-10-004-115B-34
4	84	11.1	385	7	US-60-586-841-1453
5	82.5	10.9	597	6	US-10-784-592-35
6	82	10.8	1236	6	US-10-873-528-109
7	81.5	10.8	1158	6	US-10-170-205E-32312
8	81.5	10.8	1158	6	US-10-501-035-366
9	80.5	10.6	1095	5	US-09-806-709-69
10	80.5	10.6	1535	5	US-09-806-709-67
11	80	10.6	485	6	US-10-170-205E-11455
12	80	10.6	503	6	US-10-170-205E-11130
13	80	10.6	503	6	US-10-170-205E-11456
14	80	10.6	546	6	US-10-425-115-362763
15	80	10.6	829	7	US-60-579-902-7335
16	79	10.4	4326	1	PCT-US04-13717-18
17	79	10.4	4326	6	US-10-831-070-18
18	78.5	10.4	358	6	US-10-767-701-40444
19	77	10.2	841	7	US-60-585-632-7906
20	77	10.2	841	7	US-60-579-062-7906
21	77	10.2	10203	6	US-10-724-972A-4098
22	76.5	10.1	388	6	US-10-805-394-5044
23	76.5	10.1	630	7	US-60-581-351-9806
24	76.5	10.1	631	7	US-60-581-351-9807
25	76.5	10.1	892	5	US-09-952-267B-5
26	76.5	10.1	892	6	US-10-872-768-5

27 76.5 10.1 892 6 US-10-872-769-5 Sequence 5, Appli  
28 76.5 10.1 1627 7 US-60-546-745-52 Sequence 52, Appli  
29 412 7 US-60-585-632-7905 Sequence 7905, Ap  
30 76 10.0 412 7 US-60-579-062-7905 Sequence 7905, Ap  
31 76 10.0 852 6 US-10-170-205E-3413 Sequence 3413, Ap  
32 75.5 10.0 408 5 US-09-248-796A-19565 Sequence 19565, A  
33 75.5 10.0 839 1 PCT-US04-12070-2 Sequence 2, Appli  
34 75.5 10.0 1531 1 PCT-US03-32645-44 Sequence 44, Appli  
35 75.5 10.0 1924 1 PCT-US04-12717-78 Sequence 78, Appli  
36 75.5 10.0 1924 6 US-10-831-070-78 Sequence 78, Appli  
37 75 9.9 283 7 US-60-585-632-12923 Sequence 12923, A  
38 75 9.9 283 7 US-60-579-062-12923 Sequence 12923, A  
39 75 9.9 3132 7 US-60-556-841-1245 Sequence 1245, Ap  
40 75 9.9 3469 1 PCT-US03-24982A-255 Sequence 255, Ap  
41 75 9.9 3604 6 US-10-745-237-24 Sequence 24, Appli  
42 74.5 9.8 215 5 US-09-248-796A-20202 Sequence 20202, A  
43 74.5 9.8 237 6 US-10-425-115-218336 Sequence 218336  
44 74.5 9.8 393 6 US-10-170-205E-6541 Sequence 6541, Ap  
45 74.5 9.8 444 6 US-10-425-115-330283 Sequence 330283,

## ALIGNMENTS

## RESULT 1

US-09-741-873C-4  
; Sequence 4, Application US/09741873C  
; GENERAL INFORMATION:  
; APPLICANT: Normark, Staffan  
; APPLICANT: Olsen, Arne  
; TITLE OF INVENTION: Fibronectin Binding Protein As Well As Its Preparation  
; FILE REFERENCE: 012889-084  
; CURRENT APPLICATION NUMBER: US/09/741,873C  
; CURRENT FILING DATE: 2000-12-22  
; PRIOR APPLICATION NUMBER: SE 8801723-1  
; PRIOR FILING DATE: 1998-05-06  
; PRIOR APPLICATION NUMBER: US 08/978,878  
; PRIOR FILING DATE: 1997-11-26  
; PRIOR APPLICATION NUMBER: US 07/347,189  
; PRIOR FILING DATE: 1989-05-04  
; PRIOR APPLICATION NUMBER: US 07/799,437  
; PRIOR FILING DATE: 1991-11-06  
; PRIOR APPLICATION NUMBER: US 07/970,846  
; PRIOR FILING DATE: 1992-11-03  
; PRIOR APPLICATION NUMBER: US 08/187,865  
; PRIOR FILING DATE: 1994-01-28  
; PRIOR APPLICATION NUMBER: US 08/318,519  
; PRIOR FILING DATE: 1994-10-05  
; NUMBER OF SEQ ID NOS: 11  
; SOFTWARE: PatentIn version 3.0  
; SEQ ID NO 4  
; LENGTH: 151  
; TYPE: PRT  
; ORGANISM: Escherichia coli  
US-09-741-873C-4

Query Match 63.5%; Score 481; DB 5; Length 151;  
Best Local Similarity 66.2%; Pred. No. 1.4e-39;  
Matches 100; Conservative 19; Mismatches 32; Indels 0; Gaps 0;  
QY 1 MLLKVAFAAIVVSGSALAGVVDLVTRVYVTHMAHAGSPDSTLSIYQVGSANAALAQ 60  
Db 1 MLLKVAFAAIVVSGSALAGVVDLVTRVYVTHMAHAGSPDSTLSIYQVGSANAALAQ 60  
QY 61 SDARKSETTTTQSGYNGADVGAGDNTTTLTQNGFRNATIDQWAKNSDITVQYGG 120  
Db 61 TDARNSDLTTTQGGGNGADVGAGDSDSSLDLTFQFGNSATLDQWNGKNSMTVKQFGG 120  
QY 121 NNAALVNTQASDSSVMVROVQFGNNATANOY 151  
Db 121 GNGAAVDQASNSNVTVQVGFNNATANOY 151

## RESULT 2

US-09-741-873C-2  
 ; Sequence 2, Application US/09741873C  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Normark, Staffan  
 ; APPLICANT: Olsen, Arne  
 ; TITLE OF INVENTION: Fibronectin Binding Protein As Well As Its Preparation  
 ; FILE REFERENCE: 012889-084  
 ; CURRENT APPLICATION NUMBER: US/09741,873C  
 ; CURRENT FILING DATE: 2000-12-22  
 ; PRIOR APPLICATION NUMBER: SE 8601723-1  
 ; PRIOR FILING DATE: 1998-05-06  
 ; PRIOR APPLICATION NUMBER: US 08/978,878  
 ; PRIOR FILING DATE: 1997-11-26  
 ; PRIOR APPLICATION NUMBER: US 07/347,189  
 ; PRIOR FILING DATE: 1989-05-04  
 ; PRIOR APPLICATION NUMBER: US 07/789,437  
 ; PRIOR FILING DATE: 1991-11-06  
 ; PRIOR APPLICATION NUMBER: US 07/970,846  
 ; PRIOR FILING DATE: 1992-11-03  
 ; PRIOR APPLICATION NUMBER: US 08/187,865  
 ; PRIOR FILING DATE: 1994-01-28  
 ; PRIOR APPLICATION NUMBER: US 08/318,519  
 ; PRIOR FILING DATE: 1994-10-05  
 ; NUMBER OF SEQ ID NOS: 11  
 ; SOFTWARE: Patentin version 3.0  
 ; SEQ ID NO 2  
 ; LENGTH: 131  
 ; TYPE: PRT  
 ; ORGANISM: Escherichia coli  
 US-09-741-873C-2

Query Match 54.7%; Score 414; DB 5; Length 131;  
 Best Local Similarity 59.9%; Pred. No. 4.3e-33;  
 Matches 79; Conservative 17; Mismatches 17; Indels 0; Gaps 0;  
 QY 39 SGPDSLTIYQYGSANAALQSDARKSETITQSGYNGADVGGQADNSTIELTQNGPR 98  
 DB 19 SGNSELNTYQYGGNSALALQTDARNSLTIITQHGCGNADVGGQSDSSIDLQRGFG 78  
 QY 99 NNATIDOWNAKNSDITVQYGGNNAALVNQTSADSSVMYRVQVFGNNATANYQ 151  
 DB 79 NSATLDWNGKNSMTVKQGGCGGAADVDTASNSNVVTVQVFGNNATANYQ 131

## RESULT 3

US-10-004-115B-34  
 ; Sequence 34, Application US/10004115B  
 ; GENERAL INFORMATION:  
 ; APPLICANT: ASAKO, HIROYUKI  
 ; APPLICANT: MATSUMURA, KENJI  
 ; APPLICANT: SHIMIZU MASATOSHI  
 ; APPLICANT: ITO, NOSUYA  
 ; APPLICANT: WAKITA, RYUHEI  
 ; TITLE OF INVENTION: PROCESS FOR PRODUCING OPTICALLY ACTIVE  
 ; FILE REFERENCE: 4-HALO-3-HYDROXYBUTANOATE  
 ; CURRENT APPLICATION NUMBER: US/10/004,115B  
 ; CURRENT FILING DATE: 2001-12-06  
 ; PRIOR APPLICATION NUMBER: JP 2000-372704  
 ; PRIOR FILING DATE: 2000-12-07  
 ; PRIOR APPLICATION NUMBER: JP 2001-006144  
 ; PRIOR FILING DATE: 2001-01-15  
 ; PRIOR APPLICATION NUMBER: JP 2001-026594  
 ; PRIOR FILING DATE: 2001-02-02  
 ; PRIOR APPLICATION NUMBER: JP 2001-175175  
 ; PRIOR FILING DATE: 2001-06-11  
 ; NUMBER OF SEQ ID NOS: 37  
 ; SOFTWARE: Patentin Ver. 3.2  
 ; SEQ ID NO 34  
 ; LENGTH: 348  
 ; TYPE: PRT  
 ; ORGANISM: Corynebacterium sp.

## US-10-004-115B-34

Query Match 12.3%; Score 93; DB 6; Length 348;  
 Best Local Similarity 24.8%; Pred. No. 0.43;  
 Matches 40; Conservative 26; Mismatches 65; Indels 30; Gaps 6;  
 QY 1 MKLKVAFAAIVVSGSALAGVYDQVLTVRVVTTHMAHASGPDSTLSIYQY-GSANAALAL 59  
 DB 191 LRHLSAATVIALDVSADKL-----ELATKVGAEHV-LSKDAENVRKITSGGAALVL 244  
 QY 60 QSDARKSETTITQSGYNGADV-----GGADNSTIELTQNGFRNATIDOWNAKNS --- 111  
 DB 245 DFVGYPITDTAMAVAGVGSDDVTIVIGDGOAHAKVGFQSPYEASVTVPYWGARNELIE 304  
 QY 112 -----DITVQYGGNNA-----LVNQTSADSSVMV 137  
 DB 305 LIDLAKHAGIFDISVETFSLDNGAEAYRRLAAGTILSGEAVVV 345

## RESULT 4

US-60-556-841-1453  
 ; Sequence 1453, Application US/60556841  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Abad, Mark S.  
 ; TITLE OF INVENTION: Genes and Uses for Plant Improvement  
 ; FILE REFERENCE: 38-21(53450)  
 ; CURRENT APPLICATION NUMBER: US/60/556,841  
 ; CURRENT FILING DATE: 2004-03-25  
 ; NUMBER OF SEQ ID NOS: 12463  
 ; SEQ ID NO 1453  
 ; LENGTH: 385  
 ; TYPE: PRT  
 ; ORGANISM: Corynebacterium sp. ST-10  
 US-60-556-841-1453

Query Match 11.1%; Score 84; DB 7; Length 385;  
 Best Local Similarity 25.9%; Pred. No. 3.7;  
 Matches 30; Conservative 21; Mismatches 53; Indels 12; Gaps 4;  
 QY 1 MKLKVAFAAIVVSGSALAGVYDQVLTVRVVTTHMAHASGPDSTLSIYQY-GSANAALAL 59  
 DB 191 LRHLSAATVIALDVSADKL-----ELATKVGAEHV-LSKDAENVRKITSGGAALVL 244  
 QY 60 QSDARKSETTITQSGYNGADV-----GGADNSTIELTQNGFRNATIDOWNAKN 110  
 DB 245 DFVGYPITDTAMAVAGVGSDDVTIVIGDGOAHAKVGFQSPYEASVTVPYWGARN 300

## RESULT 5

US-10-784-592-35  
 ; Sequence 35, Application US/10784592  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Wilting, Reinhard  
 ; APPLICANT: Ostergaard, Peter  
 ; APPLICANT: Lassen, Soren  
 ; TITLE OF INVENTION: POLYPEPTIDES OF ALICYCLOBACILLUS SP.  
 ; FILE REFERENCE: 10406.203-US  
 ; CURRENT APPLICATION NUMBER: US/10/784,592  
 ; CURRENT FILING DATE: 2004-02-23  
 ; NUMBER OF SEQ ID NOS: 55  
 ; SOFTWARE: Patentin version 3.2  
 ; SEQ ID NO 35  
 ; LENGTH: 597  
 ; TYPE: PRT  
 ; ORGANISM: Alicyclobacillus sp.  
 ; FEATURE:  
 ; NAME/KEY: SIGNAL  
 ; LOCATION: (1)..(49)  
 ; FEATURE:  
 ; NAME/KEY: mat peptide  
 ; LOCATION: (50)..(597)  
 ; OTHER INFORMATION: multi copper oxidase  
 ; FEATURE:

NAME/KEY: MISC\_FEATURE  
LOCATION: (139)..(139)  
OTHER INFORMATION: putative copper binding site  
FEATURE:  
NAME/KEY: MISC\_FEATURE  
LOCATION: (141)..(141)  
OTHER INFORMATION: putative copper binding site  
FEATURE:  
NAME/KEY: MISC\_FEATURE  
LOCATION: (181)..(181)  
OTHER INFORMATION: putative copper binding site  
FEATURE:  
NAME/KEY: MISC\_FEATURE  
LOCATION: (183)..(183)  
OTHER INFORMATION: putative copper binding site  
FEATURE:  
NAME/KEY: MISC\_FEATURE  
LOCATION: (514)..(514)  
OTHER INFORMATION: putative copper binding site  
FEATURE:  
NAME/KEY: MISC\_FEATURE  
LOCATION: (566)..(566)  
OTHER INFORMATION: putative copper binding site  
US-10-784-592-35

Query Match 10.9%; Score 82.5; DB 6; Length 597;  
Best Local Similarity 26.1%; Pred. No. 9;  
Matches 46; Conservative 27; Mismatches 50; Indels 53; Gaps 11;

QY 7 AAFPAIVVSGSALAGVYDQVTRVVTHEMAHAGSPDSTLSIYQYGSANAALALQSDAR-- 64

DB 185 ASFEV---GLGLGAFVLPKRAV-----HPADRDYTLVHEWPTASTAQTMWANKAG 236

QY 65 -----KSETT-----ITQSGYNG-----AD-VGQGDNST-----IELTQNGF 97

DB 237 NLGSAKESAGMGGMQNGQNDGMMGMAADCTGCGGNSASDAHVLPGPPLQINGF 296

QY 98 RNNATIDOWNAKNSDITVQYGGNNAALVNOTASDSS-----VMRQVGFQGN 144

DB 297 --SPTANDWAA--LDENAGMY---DAFTVQNGASGTTLLPAKFGQLVRLIVNSGN 345

## RESULT 6

US-10-873-528-109  
Sequence 109, Application US/10873528  
GENERAL INFORMATION:  
APPLICANT: Microbial Technics Limited  
APPLICANT: Gilbert, Christophe FG  
APPLICANT: Hansbro, Philip M  
TITLE OF INVENTION: Proteins  
FILE REFERENCE: PWC/P21129WO  
CURRENT APPLICATION NUMBER: US/10/873,528  
CURRENT FILING DATE: 2004-06-23  
PRIOR APPLICATION NUMBER: US/09/769,787  
PRIOR FILING DATE: 2001-01-26  
PRIOR APPLICATION NUMBER: GB 9816337.1  
PRIOR FILING DATE: 1998-03-27  
PRIOR APPLICATION NUMBER: US 60/125164  
PRIOR FILING DATE: 1999-03-19  
NUMBER OF SEQ ID NOS: 388  
SOFTWARE: Patent in Ver. 2.1  
SEQ ID NO 109  
LENGTH: 1236  
TYPE: PRT  
ORGANISM: Streptococcus pneumoniae  
US-10-873-528-109

Query Match 10.8%; Score 82; DB 6; Length 1236;  
Best Local Similarity 26.0%; Pred. No. 25;  
Matches 39; Conservative 16; Mismatches 61; Indels 34; Gaps 7;

QY 16 GSALAGVYDQVTRVVTHEMAHAGSPDSTLSIYQYGSANAALALQSDARK----- 65

DB 235 GSKLTFTYTVVNPKNLDLGNISSMRPGYSIYNSGTSTQTMULTGLSDLGKPSGVKNYIT 294

QY 66 -----SETTITQ-SQY--GNGADV-----GQGDNS-TIELTQNGFRNNATI 103

DB 295 DNGRQVLSTNTSTVTTQSGSYTWNGAQMGNGFFAKGYGLTSSWVPII--GTDISFTF 352

QY 104 DOWNAKNSDITVQYGGNNAALVNOTASDS 133

DB 353 TPYAARTDRIGINVFNGGKVVESSTTSQS 382

## RESULT 7

US-10-170-205E-32312  
Sequence 32312, Application US/10170205E  
GENERAL INFORMATION:  
APPLICANT: ADAMS, Mark  
TITLE OF INVENTION: DEVICES, SUCH AS ARRAYS, COMPRISED OF HUMAN PROTEINS OR PROTEIN  
CAPTURE AGENTS, AND USES THEREOF  
FILE REFERENCE: CL001381  
CURRENT APPLICATION NUMBER: US/10/170,205E  
CURRENT FILING DATE: 2002-06-13  
NUMBER OF SEQ ID NOS: 40312  
SOFTWARE: Patent in version 3.2  
SEQ ID NO 32312  
LENGTH: 1158  
TYPE: PRT  
ORGANISM: Homo sapiens  
US-10-170-205E-32312

Query Match 10.8%; Score 81.5; DB 6; Length 1158;  
Best Local Similarity 21.1%; Pred. No. 26;  
Matches 30; Conservative 28; Mismatches 67; Indels 17; Gaps 3;

QY 5 KVAFAAIVVSGSALAGVYDQVTRVVTHEMAHAGSPDSTLSIYQYGSANAALALQSDAR 64

DB 297 RLAAVARLOENGQKDVGSY-----QLPKGMSSHLNGQARTSSSKLMASKSSATVFQPMG 351

QY 65 KSETTITQSGYNGAD---VGQGDNS-----TIELTQNGFRNNATIDOWNAKNSD 112

DB 352 IIPSPKNAGYKNSLERNNIKQAANNSLLHLKLSQTIKPMNGHSHSRGSIFFESSTP 411

QY 113 ITVQYGGNNAALVNOTASDSS 134

DB 412 TTIDEYSDNNPFTDDSSGDES 433

## RESULT 8

US-10-501-035-366  
Sequence 366, Application US/10501035  
GENERAL INFORMATION:  
APPLICANT: Bristol-Myers Squibb Company  
TITLE OF INVENTION: IDENTIFICATION OF POLYNUCLEOTIDES AND POLYPEPTIDE FOR PREDICTING  
ACTIVITY OF COMPOUNDS THAT INTERACT WITH PROTEIN TYROSINE KINASE PATHWAYS  
FILE REFERENCE: D0185 PCT  
CURRENT APPLICATION NUMBER: US/10/501,035  
CURRENT FILING DATE: 2004-07-09  
PRIOR APPLICATION NUMBER: US 60/350,061  
PRIOR FILING DATE: 2002-01-18  
NUMBER OF SEQ ID NOS: 795  
SOFTWARE: Patent in version 3.2  
SEQ ID NO 366  
LENGTH: 1158  
TYPE: PRT  
ORGANISM: Homo sapiens  
US-10-501-035-366

Query Match 10.8%; Score 81.5; DB 6; Length 1158;  
Best Local Similarity 21.1%; Pred. No. 26;  
Matches 30; Conservative 28; Mismatches 67; Indels 17; Gaps 3;

QY 5 KVAFAAIVVSGSALAGVYDQVTRVVTHEMAHAGSPDSTLSIYQYGSANAALALQSDAR 64

Db 297 RLAMARLEQNGQKDVSY-----QLPKMGSHLNGQARTSSSKLMAKSSATVFQNPWG 351  
Qy 65 KSETTITQSGVGNAD---VQGGADNS-----TIELTQNGFRNNATIDOWNAKNSD 112  
Db 352 IIPSPPNAGYKNSLERNIKQARNSLLHLKLSQTIKPKPMNGHSHSERSGIFESSTP 411  
Qy 113 ITVGQYGNNAALVNOTASDSS 134  
Db 412 TTIDEYSDNAPSFDDSSGDES 433

RESULT 9  
US-09-806-709-69  
; Sequence 69, Application US/09806709  
; GENERAL INFORMATION:  
; APPLICANT: LOOSMORE, Sheena M.  
; APPLICANT: YANG, Yan-Ping  
; APPLICANT: KLEIN, Michel H.  
; TITLE OF INVENTION: PROTECTIVE RECOMBINANT HAEMOPHILUS INFLUENZAE HIGH  
; TITLE OF INVENTION: MOLECULAR WEIGHT PROTEINS  
; FILE REFERENCE: 1038-1138 MIS  
; CURRENT APPLICATION NUMBER: US/09/806,709  
; CURRENT FILING DATE: 2001-04-04  
; PRIOR APPLICATION NUMBER: PCT/CA99/00938  
; PRIOR FILING DATE: 1999-10-07  
; PRIOR APPLICATION NUMBER: 09/167,568  
; PRIOR FILING DATE: 1998-10-07  
; PRIOR APPLICATION NUMBER: 09/206,942  
; PRIOR FILING DATE: 1999-12-08  
; NUMBER OF SEQ ID NOS: 95  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 69  
; LENGTH: 1095  
; TYPE: PRT  
; ORGANISM: Haemophilus influenzae  
US-09-806-709-69

Query Match 10.6%; Score 80.5; DB 5; Length 1095;  
Best Local Similarity 26.4%; Pred. No. 30;  
Matches 39; Conservative 22; Mismatches 60; Indels 27; Gaps 7;

Qy 15 SGSALAGVYDQ--LVTRVWTH---EMAHASGPDSTLSIYQYGSANAALQSQ-DARKSET 68  
Db 727 TGSILGGIESSSGSVTLTATEGALAVNSISG--NTVTVTANSGLTTLGASTIKGTESVT 784  
Qy 69 TITQSGYNGADVGQADNSTIEL-----TQNGFRNNATIDOWNAKNSDITVQYGGN 121  
Db 785 TSSQSG-----DIGGTISGGTVEVKATESLTQNSKIKATTEANVTSGTIGTIGTISG 839  
Qy 122 NAALVNOTASDSSVMVQVGFNNATAN 149  
Db 840 NTNVNTANAGDLTV-----GNGAEIN 860

RESULT 10  
US-09-806-709-67  
; Sequence 67, Application US/09806709  
; GENERAL INFORMATION:  
; APPLICANT: LOOSMORE, Sheena M.  
; APPLICANT: YANG, Yan-Ping  
; APPLICANT: KLEIN, Michel H.  
; TITLE OF INVENTION: PROTECTIVE RECOMBINANT HAEMOPHILUS INFLUENZAE HIGH  
; TITLE OF INVENTION: MOLECULAR WEIGHT PROTEINS  
; FILE REFERENCE: 1038-1138 MIS  
; CURRENT APPLICATION NUMBER: US/09/806,709  
; CURRENT FILING DATE: 2001-04-04  
; PRIOR APPLICATION NUMBER: PCT/CA99/00938  
; PRIOR FILING DATE: 1999-10-07  
; PRIOR APPLICATION NUMBER: 09/167,568  
; PRIOR FILING DATE: 1998-10-07  
; PRIOR APPLICATION NUMBER: 09/206,942  
; PRIOR FILING DATE: 1999-12-08  
; NUMBER OF SEQ ID NOS: 95

; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 67  
; LENGTH: 1536  
; TYPE: PRT  
; ORGANISM: Haemophilus influenzae  
US-09-806-709-67

Query Match 10.6%; Score 80.5; DB 5; Length 1536;  
Best Local Similarity 26.4%; Pred. No. 46;  
Matches 39; Conservative 22; Mismatches 60; Indels 27; Gaps 7;

Qy 15 SGSALAGVYDQ--LVTRVWTH---EMAHASGPDSTLSIYQYGSANAALQSQ-DARKSET 68  
Db 1168 TGSILGGIESSSGSVTLTATEGALAVNSISG--NTVTVTANSGLTTLGASTIKGTESVT 1225  
Qy 69 TITQSGYNGADVGQADNSTIEL-----TQNGFRNNATIDOWNAKNSDITVQYGGN 121  
Db 1226 TSSQSG-----DIGGTISGGTVEVKATESLTQNSKIKATTEANVTSGTIGTIGTISG 1280  
Qy 122 NAALVNOTASDSSVMVQVGFNNATAN 149  
Db 1281 NTNVNTANAGDLTV-----GNGAEIN 1301

RESULT 11  
US-10-170-205E-11455  
; Sequence 11455, Application US/10170205E  
; GENERAL INFORMATION:  
; APPLICANT: ADAMS, Mark  
; TITLE OF INVENTION: DEVICES, SUCH AS ARRAYS, COMPRISED OF HUMAN PROTEINS OR PROTEIN  
; TITLE OF INVENTION: CAPTURE AGENTS, AND USES THEREOF  
; FILE REFERENCE: CLO01381  
; CURRENT APPLICATION NUMBER: US/10/170,205E  
; CURRENT FILING DATE: 2002-06-13  
; NUMBER OF SEQ ID NOS: 40312  
; SOFTWARE: PatentIn version 3.2  
; SEQ ID NO 11455  
; LENGTH: 485  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-10-170-205E-11455

Query Match 10.6%; Score 80; DB 6; Length 485;  
Best Local Similarity 28.2%; Pred. No. 12;  
Matches 35; Conservative 14; Mismatches 53; Indels 22; Gaps 5;

Qy 32 THEMAHA-----SGPDSTLSIYQYGSANAALQSQDARKSETTIT--OSGYGNGA 79  
Db 276 THSTAHLPVQTVQTFESNPDGTVSLIOVGT-GATVATLADASELPTTVTAQVNVSAVA 334  
Qy 80 DVQGGADN-----STIELTQNGFRNNATIDOWNAKNSDITVQYGGNNAALVNOTAS 131  
Db 335 D-GEVEQNWATLQGGEMTITQTQASEATQAVASLAEEAAVAASQEMQOGATVTTMALNSEAA 393  
Qy 132 DSSV 135  
Db 394 AHAV 397

RESULT 12  
US-10-170-205E-11130  
; Sequence 11130, Application US/10170205E  
; GENERAL INFORMATION:  
; APPLICANT: ADAMS, Mark  
; TITLE OF INVENTION: DEVICES, SUCH AS ARRAYS, COMPRISED OF HUMAN PROTEINS OR PROTEIN  
; TITLE OF INVENTION: CAPTURE AGENTS, AND USES THEREOF  
; FILE REFERENCE: CLO01381  
; CURRENT APPLICATION NUMBER: US/10/170,205E  
; CURRENT FILING DATE: 2002-06-13  
; NUMBER OF SEQ ID NOS: 40312  
; SOFTWARE: PatentIn version 3.2  
; SEQ ID NO 11130  
; LENGTH: 503

```
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-170-205E-11130

Query Match      10.6%; Score 80; DB 6; Length 503;
Best Local Similarity 28.2%; Pred. No. 13;
Matches 35; Conservative 14; Mismatches 53; Indels 22; Gaps 5;

QY 32 THEMALIA-----SGPDSTLSIYQYGSANAALALQSDARKSETTIT--QSGYGNGA 79
Db 294 THSIAHLVPSQTVVQTFSPDGVSLIQVGT-GATVATLADASELPTTIVTAQVNYSAVA 352
QY 80 DVGGADN-----STIELTQNGFRNNATIDQWNAKNSDITVGQYGGNNAALVNQTA 131
Db 353 D-GEVEQNWATLOGGEMTIGTQASEATQAVASLAEEAASQEMQOGATVTMALNSEAA 411
QY 132 DSSV 135
Db 412 AHAV 415

RESULT 13
US-10-170-205E-11456
; Sequence 11456, Application US/10170205E
; GENERAL INFORMATION:
; APPLICANT: ADAMS, Mark
; TITLE OF INVENTION: DEVICES, SUCH AS ARRAYS, COMPRISED OF HUMAN PROTEINS OR PROTEIN
; TITLE OF INVENTION: CAPTURE AGENTS, AND USES THEREOF
; FILE REFERENCE: CL001381
; CURRENT APPLICATION NUMBER: US/10/170,205E
; CURRENT FILING DATE: 2002-06-13
; NUMBER OF SEQ ID NOS: 40312
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 11456
; LENGTH: 503
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-170-205E-11456

Query Match      10.6%; Score 80; DB 6; Length 503;
Best Local Similarity 28.2%; Pred. No. 13;
Matches 35; Conservative 14; Mismatches 53; Indels 22; Gaps 5;

QY 32 THEMALIA-----SGPDSTLSIYQYGSANAALALQSDARKSETTIT--QSGYGNGA 79
Db 294 THSIAHLVPSQTVVQTFSPDGVSLIQVGT-GATVATLADASELPTTIVTAQVNYSAVA 352
QY 80 DVGGADN-----STIELTQNGFRNNATIDQWNAKNSDITVGQYGGNNAALVNQTA 131
Db 353 D-GEVEQNWATLOGGEMTIGTQASEATQAVASLAEEAASQEMQOGATVTMALNSEAA 411
QY 132 DSSV 135
Db 412 AHAV 415

RESULT 14
US-10-425-115-362763
; Sequence 362763, Application US/10425115
; GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J.
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
; TITLE OF INVENTION: Plants
; FILE REFERENCE: 38-21(53222)B
; CURRENT APPLICATION NUMBER: US/10/425,115
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 369326
; SEQ ID NO 362763
; LENGTH: 546
; TYPE: PRT
```

```
; ORGANISM: Zea mays
; FEATURE:
; NAME/KEY: unsure
; LOCATION: (1)..(546)
; OTHER INFORMATION: unsure at all Xaa locations
; FEATURE:
; OTHER INFORMATION: Clone ID: MRT4577_94013C.1.pcp
US-10-425-115-362763

Query Match      10.6%; Score 80; DB 6; Length 546;
Best Local Similarity 27.2%; Pred. No. 14;
Matches 34; Conservative 14; Mismatches 43; Indels 34; Gaps 5;

QY 1 MKLLKVAFAAIVVSGSALAGVYDQLVTRVVTTHMAHASGPDSTLS-IYQYGSANAALAL 59
Db 90 LMLRLGAVPLVWSSPRAA-----EATR--THDHVFASRPTSTLSDEILYSSDIAFSP 143
QY 60 QSD-----ARKESETTITGSGYNGADVGCGANSTIELTQN 95
Db 144 YGEHWRQARKLVTAHLFTVKRVHSYRRARKEEVLVLVAKVREAAVAGTATDMS---LAMN 200
QY 96 GFRNN 100
Db 201 TFAND 205

RESULT 15
US-60-579-902-7335
; Sequence 7335, Application US/60579902
; GENERAL INFORMATION:
; APPLICANT: Goldman, Barry S.
; APPLICANT: Krasomil-Osterfeld, Karina C.
; APPLICANT: Wu, Wei
; TITLE OF INVENTION: Nucleotide and Amino Acid Sequences from Xenorhabdus
; TITLE OF INVENTION: and Uses thereof
; FILE REFERENCE: 38-21 (53444) A
; CURRENT APPLICATION NUMBER: US/60/579,902
; CURRENT FILING DATE: 2004-06-15
; NUMBER OF SEQ ID NOS: 14985
; SEQ ID NO 7335
; LENGTH: 829
; TYPE: PRT
; ORGANISM: Xenorhabdus bovienii
US-60-579-902-7335

Query Match      10.6%; Score 80; DB 7; Length 829;
Best Local Similarity 25.2%; Pred. No. 24;
Matches 36; Conservative 24; Mismatches 53; Indels 30; Gaps 6;

QY 23 YDQVTRVVTTHMAHASGPDSTLSIYQYGSANAALALQSD-ARKSETTITGSGY-NGAD 80
Db 474 HDLE-----SYQKHKFQNLISQNLSDYSGIYVSGYLQDDWGSRTTSLNVGYSVNYAD 528
QY 81 VGCGADNSTIELTQNGFRNNATIDQWNAKNSDITVGQY-----GGNNA 124
Db 529 I-----NYTLNVTQNWSSGNSTADNQLAFNIQVPLDRWLPNSWATYSLNHNRRQGMNQSA 583
QY 125 LVNQTASDSSVM--VRQVGFQNN 145
Db 584 GLSGTALDGNVLGYSLQORYGNN 606

Search completed: August 2, 2004, 15:29:52
Job time : 18.8 secs
```





GenCore version 5.1.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: August 2, 2004, 14:39:53 ; Search time 9.4 Seconds  
(without alignments)  
1545.204 Million cell updates/sec

Title: US-09-543-407-16  
Perfect score: 757  
Sequence: 1 MKLLKVAFAAIVSSSALA.....DSSVMVRQVGFNNATANQY 151

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 283366 seqs, 96191526 residues

Total number of hits satisfying chosen parameters: 283366

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : PIR 78:\*  
1: pir1:\*  
2: pir2:\*  
3: pir3:\*  
4: pir4:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	658	86.9	151	2 JC6039	fimbrin protein ag
2	658	86.9	151	2 A10635	major curlin chain
3	484	63.9	151	2 S70788	curlin protein csg
4	474.5	62.7	152	2 D90806	curlin major subun
5	474.5	62.7	152	2 H85665	hypothetical prote
6	118.5	15.7	151	2 S70787	curlin nucleator p
7	118.5	15.7	151	2 C90806	minor curlin subun
8	118.5	15.7	151	2 G85665	curlin minor chain
9	111.5	14.7	151	2 JC6040	fimbrin protein ag
10	111.5	14.7	151	2 AH0635	nucleation compone
11	101.5	13.4	590	1 A45621	leishmanolysin (EC
12	101	13.3	590	2 B42049	leishmanolysin (EC
13	101	13.3	593	2 A44951	probable PPE prote
14	99.5	13.1	582	2 F70675	leishmanolysin (EC
15	99.5	13.1	646	1 S19916	leishmanolysin (EC
16	98	12.9	145	2 AD3143	conserved hypothet
17	98	12.9	145	2 H98144	hypothetical prote
18	97.5	12.9	1034	2 JC2143	ice nucleation act
19	96.5	12.7	602	1 PL0221	leishmanolysin (EC
20	96	12.7	1651	2 JC1340	outer membrane pro
21	95.5	12.6	1258	2 JC0188	ice nucleation pro
22	94.5	12.5	1322	2 S07053	ice nucleation pro
23	93.5	12.4	1567	2 S11672	ice nucleation prote
24	91.5	12.1	1655	2 E97835	hypothetical prote
25	89.5	11.8	552	2 D70604	probable PPE prote
26	89.5	11.8	639	2 C42049	leishmanolysin (EC
27	88.5	11.7	590	2 E70946	probable PPE prote
28	87	11.5	329	2 S23247	outer membrane pro
29	87	11.5	331	2 S21406	outer membrane pro

30	87	11.5	331	2 S21408	outer membrane pro
31	87	11.5	455	2 C29349	hypothetical prote
32	87	11.5	573	2 C86266	R3F19_21 protein -
33	86.5	11.4	3300	2 D70575	probable PPE prote
34	85.5	11.3	1210	2 A25547	ice nucleation pro
35	84	11.1	823	2 S14055	nucleoskeletal-lik
36	84	11.1	3716	2 E70969	probable PPE prote
37	83.5	11.0	1200	1 SNPSO	ice nucleation pro
38	82	10.8	528	2 S69589	hypothetical prote
39	82	10.8	534	2 T32020	hypothetical prote
40	82	10.8	1026	2 A48995	paracrystalline su
41	82	10.8	1073	2 C87374	S-layer protein R8
42	82	10.8	1635	2 A10452	hemolysin limpor
43	82	10.8	4776	2 E95206	cell wall surface
44	81.5	10.8	760	2 C90739	hypothetical prote
45	81.5	10.8	760	2 E85589	hypothetical prote

ALIGNMENTS

RESULT 1

JC6039  
fimbrin protein agfa precursor - Salmonella enteritidis  
C:Species: Salmonella enteritidis  
C:Date: 31-Dec-1996 #sequence\_revision 31-Dec-1996 #text\_change 08-Oct-1999  
C:Accession: JC6039; PC6015; A44898  
R:Collinson, S.K.; Clouthier, S.C.; Doran, J.L.; Baner, P.A.; Kay, W.W.  
J. Bacteriol. 178, 662-667, 1996  
A:Title: Salmonella enteritidis agfBAC operon encoding thin, aggregative fimbriae.  
A:Reference number: JC6039; MUID:96146512; PMID:8550497  
A:Accession: JC6039  
A:Molecule type: DNA  
A:Residues: 1-151 <COL>  
A:Cross-references: GB:U43280; NID:gl184712; PIDN:AAC43599.1; PID:gl184714  
A:Accession: PC6015  
A:Molecule type: Protein  
A:Residues: 21-52 <CO2>  
A:Experimental source: strain 27655-3b  
A:Note: the authors translated the codon ACG for residue 44 as Ile  
R:Collinson, S.K.; Emody, L.; Muller, K.H.; Trust, T.J.; Kay, W.W.  
J. Bacteriol. 173, 4773-4781, 1991  
A:Title: Purification and characterization of thin, aggregative fimbriae from Salmonella  
A:Reference number: A44898; MUID:91310586; PMID:1677357  
A:Contents: 27655  
A:Accession: A44898  
A>Status: preliminary  
A:Molecule type: protein  
A:Residues: 21-33 <CO3>  
A:Note: sequence extracted from NCBI backbone (NCBIP:45936)  
C:Genetics:  
A:Gene: agfa  
C:Function:  
A:Description: major component of thin aggregative fimbriae  
A:Note: fimbriae bind to fibronectin, plasminogen, tissue plasminogen activator  
C:Keywords: fimbria  
F:1-20/Domain: signal sequence #status predicted <SIG>  
F:21-151/Product: fimbrin protein agfa #status experimental <WAT>

Query Match 86.9%; Score 658; DB 2; Length 151;  
Best Local Similarity 90.1%; Pred. No. 7.5e-52;  
Matches 136; Conservative 3; Mismatches 12; Indels 0; Gaps 0;

QY	1	MKLLKVAFAAIVSSSALAGVYDQLVTRVVTHEMAHASGPDSTLSIYQGSANAALALQ	60
DB	1	MKLLKVAFAAIVSSSALAGVFPWGCGGNHNGGSSGPDSTLSIYQGSANAALALQ	60
QY	61	SDARKSETTITQSGYNGADYVQGGADNSTIELTQGFNNATIDOWNAKNSDITVQYGG	120
DB	61	SDARKSETTITQSGYNGADYVQGGADNSTIELTQGFNNATIDOWNAKNSDITVQYGG	120
QY	121	NNAALVNQTADSSVMVRQVGFNNATANQY	151



A:Experimental source: strain K-12, substrain MG1655  
C:Genetics: CsgB  
A:Gene: csgB  
A:Map position: 23.15  
C:Function:  
A:Description: minor component of wild-type curli; interaction between CsgA and CsgB tri-  
A:Note: curli are thin, coiled fibers expressed on the surface of *Escherichia coli* that i-  
and H-kinogen; in the absence of CsgA, CsgB can self-assemble into polymers  
F:1-21/Domain: signal sequence #status predicted <SIG>  
F:22-151/Product: minor curlin chain #status predicted <MAT>

Query Match 15.7%; Score 118.5; DB 2; Length 151;  
Best Local Similarity 31.6%; Pred. No. 0.0014;  
Matches 37; Conservative 13; Mismatches 62; Indels 5; Gaps 2;

Qy 35 MAHASGPDSTLSIYQYGSAALALQSDARKSETTITQSGYNGADVGGADNSTIELTQ 94  
Db 18 IAAAGYDLANSEYNF-----AVNELSKSFNQAAIIGQAGTNNSAQLROGSGKLLAVAAQ 73  
Qy 95 NGFRNNATIDQWNAKNSDITVQYGGNNAALVNQTASDSSVMVRQVGFNGNATANQY 151  
Db 74 EGSSNRAKIDTQDYNL-AYIDQAGSANDASISQAGYNTAMIIQKSGSKANITQY 129

RESULT 7  
C90806  
minor curlin subunit precursor CsgB [imported] - *Escherichia coli* (strain O157:H7, sub-  
C:Species: *Escherichia coli*  
C:Date: 18-Jul-2001 #sequence\_revision 18-Jul-2001 #text\_change 18-Jul-2001  
C:Accession: C90806  
F:Hayashi, T.; Makino, K.; Ohnishi, M.; Kurokawa, K.; Ishii, K.; Yokoyama, K.; Han, C.G.;  
gasawara, N.; Yasunaga, T.; Kuhara, S.; Shiba, T.; Hattori, M.; Shinagawa, H.  
DNA Res. 8, 11-22, 2001  
A:Title: Complete genome sequence of enterohemorrhagic *Escherichia coli* O157:H7 and gen-  
A:Reference number: A99629; MUID:21156231; PMID:11258796  
A:Accession: C90806  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-151 <HAY>  
A:Cross-references: GB:BA000007; PIDN:BA034842.1; PID:G13360879; GSPDB:GN00154  
A:Experimental source: strain O157:H7, substrain RIMD 0509952  
C:Genetics:  
A:Gene: ECs1419

Query Match 15.7%; Score 118.5; DB 2; Length 151;  
Best Local Similarity 31.6%; Pred. No. 0.0014;  
Matches 37; Conservative 13; Mismatches 62; Indels 5; Gaps 2;

Qy 35 MAHASGPDSTLSIYQYGSAALALQSDARKSETTITQSGYNGADVGGADNSTIELTQ 94  
Db 18 IAAAGYDLANSEYNF-----AVNELSKSFNQAAIIGQAGTNNSAQLROGSGKLLAVAAQ 73  
Qy 95 NGFRNNATIDQWNAKNSDITVQYGGNNAALVNQTASDSSVMVRQVGFNGNATANQY 151  
Db 74 EGSSNRAKIDTQDYNL-AYIDQAGSANDASISQAGYNTAMIIQKSGSKANITQY 129

RESULT 8  
G85665  
curlin minor chain precursor, CsgA homolog [imported] - *Escherichia coli* (strain O157:H7;  
C:Species: *Escherichia coli*  
C:Date: 16-Feb-2001 #sequence\_revision 16-Feb-2001 #text\_change 14-Sep-2001  
C:Accession: G85665  
R:Perona, N.R.; Plunkett III, G.; Burland, V.; Mau, B.; Glasner, J.D.; Rose, D.J.; Mayhew,  
filler, L.; Grobeck, E.J.; Davis, N.W.; Lam, A.; Dimalanta, E.; Potamouis, K.; Apodaca,  
Nature 409, 529-533, 2001  
A:Title: Genome sequence of enterohemorrhagic *Escherichia coli* O157:H7.  
A:Reference number: A85480; MUID:21074935; PMID:11206551  
A:Accession: G85665  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-151 <STO>  
A:Cross-references: GB:AE005174; NID:G12514573; PIDN:AA055787.1; GSPDB:GN00145; UWPG:Z167



R;Ramamoorthy, R.; Donelson, J.E.; Paetz, K.E.; Maybodi, M.; Roberts, S.C.; Wilson, M.E.  
J. Biol. Chem. 267, 1888-1895, 1992  
A;Title: Three distinct RNAs for the surface protease gp63 are differentially expressed  
A;Reference number: A42049; MUID:92112918; PMID:1370484  
A;Accession: B42049  
A;Status: preliminary  
A;Molecule type: mRNA  
A;Residues: 1-599 <RAM>  
A;Cross-references: GB:M80669; NID:gl59324; PIDN:AAA29236.1; PID:gl59325  
C;Function:  
A;Description: catalyzes the hydrolysis of peptide bonds between two hydrophobic residue  
A;Note: the activated form can activate the proenzyme form  
C;Superfamily: leishmanolysin  
C;Keywords: cell adhesion; glycoprotein; hydrolase; membrane bound; metalloproteinase; z  
F;1-39/Domain: signal sequence #status predicted <SIG>  
F;40-97/Domain: activation peptide #status predicted <ATP>  
F;98-599/Product: leishmanolysin #status predicted <ATP>  
F;48,261,265,331/Binding site: zinc, catalytic (Cys, His, His, His) (inhibited) #status  
F;122-139,188-227,311-383,390-451,403-422,412-486,463-507,512-562,532-555/Disulfide bond  
F;261,265,331/Binding site: zinc, catalytic (His) (active) #status predicted  
F;262/Active site: Glu #status predicted  
F;297,394/Binding site: carboxylate (Asn) (covalent) #status predicted  
C;Function:  
Query Match 13.3%; Score 101; DB 2; Length 599;  
Best Local Similarity 30.2%; Pred. No. 0.27;  
Matches 32; Conservative 12; Mismatches 34; Indels 28; Gaps 3;  
QY 8 AFAIVVSGSALAGVYDQVLTTRVVTTHMAHAGSDPTLSIYQYGSANAALQSDAR--- 64  
DB 236 AVGVINIPANIASRYDQVLTTRVVTTHMAHALG-----FSVGFEGARILESINVRH 288  
QY 65 -----KSETTITQSGYNGADV-----GQADNSTIEL 92  
DB 289 KDFVPVINSSTAVAKAREQYCDTLEILEIDQGGAGSAGSHIKM 334  
RESULT 13  
A;Residues: 1-582 <COL>  
A;Cross-references: GB:M28527; NID:gl59322; PIDN:AAA29235.1; PID:gl59323  
R;Ramamoorthy, R.; Donelson, J.E.; Paetz, K.E.; Maybodi, M.; Roberts, S.C.; Wilson, M.E.  
J. Biol. Chem. 267, 1888-1895, 1992  
A;Title: Three distinct RNAs for the surface protease gp63 are differentially expressed  
A;Reference number: A42049; MUID:90205976; PMID:2320059  
A;Accession: A44951  
A;Status: preliminary  
A;Molecule type: DNA  
A;Residues: 1-599 <MIL>  
A;Cross-references: GB:M28527; NID:gl59322; PIDN:AAA29235.1; PID:gl59323  
R;Ramamoorthy, R.; Donelson, J.E.; Paetz, K.E.; Maybodi, M.; Roberts, S.C.; Wilson, M.E.  
J. Biol. Chem. 267, 1888-1895, 1992  
A;Title: Three distinct RNAs for the surface protease gp63 are differentially expressed  
A;Reference number: A42049; MUID:92112918; PMID:1370484  
A;Accession: A42049  
A;Status: preliminary  
A;Molecule type: mRNA  
A;Residues: 1-599 <RAM>  
A;Cross-references: GB:M80672; NID:gl59328; PIDN:AAA29238.1; PID:gl59329  
A;Note: sequence extracted from NCBI backbone (NCBIN:76040, NCBIIP:76041)  
A;Note: the source is designated as Leishmania donovani chagasi  
C;Function:  
A;Description: catalyzes the hydrolysis of peptide bonds between two hydrophobic residue  
A;Note: the activated form can activate the proenzyme form  
C;Superfamily: leishmanolysin  
C;Keywords: blocked carboxyl end; cell adhesion; glycoprotein; hydrolase; lipoprotein; m  
F;1-39/Domain: signal sequence #status predicted <SIG>  
F;40-97/Domain: activation peptide #status predicted <ATP>  
F;98-599/Product: leishmanolysin #status predicted <ATP>  
F;48,261,265,331/Binding site: zinc, catalytic (Cys, His, His, His) (inhibited) #status

F;122-139,188-227,311-383,390-451,403-422,412-486,463-507,512-562,532-555/Disulfide bonds  
F;261,265,331/Binding site: zinc, catalytic (His) (active) #status predicted  
F;262/Active site: Glu #status predicted  
F;394/Binding site: carboxylate (Asn) (covalent) #status predicted  
F;574/Modified site: GPI-anchor ethanolamine amidated carboxyl end (Asn) (in mature form)  
Query Match 13.3%; Score 101; DB 2; Length 599;  
Best Local Similarity 30.2%; Pred. No. 0.27;  
Matches 32; Conservative 12; Mismatches 34; Indels 28; Gaps 3;  
QY 8 AFAIVVSGSALAGVYDQVLTTRVVTTHMAHAGSDPTLSIYQYGSANAALQSDAR--- 64  
DB 236 AVGVINIPANIASRYDQVLTTRVVTTHMAHALG-----FSVGFEGARILESINVRH 288  
QY 65 -----KSETTITQSGYNGADV-----GQADNSTIEL 92  
DB 289 KDFVPVINSSTAVAKAREQYCDTLEILEIDQGGAGSAGSHIKM 334  
RESULT 14  
A;Residues: 1-582 <COL>  
A;Cross-references: GB:M28527; NID:gl59322; PIDN:AAA29235.1; PID:gl59323  
R;Ramamoorthy, R.; Donelson, J.E.; Paetz, K.E.; Maybodi, M.; Roberts, S.C.; Wilson, M.E.  
J. Biol. Chem. 267, 1888-1895, 1992  
A;Title: Three distinct RNAs for the surface protease gp63 are differentially expressed  
A;Reference number: A42049; MUID:90205976; PMID:2320059  
A;Accession: A44951  
A;Status: preliminary  
A;Molecule type: DNA  
A;Residues: 1-582 <COL>  
A;Cross-references: GB:M28527; NID:gl59322; PIDN:AAA29235.1; PID:gl59323  
R;Ramamoorthy, R.; Donelson, J.E.; Paetz, K.E.; Maybodi, M.; Roberts, S.C.; Wilson, M.E.  
J. Biol. Chem. 267, 1888-1895, 1992  
A;Title: Three distinct RNAs for the surface protease gp63 are differentially expressed  
A;Reference number: A42049; MUID:92112918; PMID:1370484  
A;Accession: A42049  
A;Status: preliminary  
A;Molecule type: mRNA  
A;Residues: 1-599 <RAM>  
A;Cross-references: GB:M80669; NID:gl59324; PIDN:AAA29236.1; PID:gl59325  
C;Function:  
A;Description: catalyzes the hydrolysis of peptide bonds between two hydrophobic residue  
A;Note: the activated form can activate the proenzyme form  
C;Superfamily: leishmanolysin  
C;Keywords: cell adhesion; glycoprotein; hydrolase; membrane bound; metalloproteinase; z  
F;1-39/Domain: signal sequence #status predicted <SIG>  
F;40-97/Domain: activation peptide #status predicted <ATP>  
F;98-599/Product: leishmanolysin #status predicted <ATP>  
F;48,261,265,331/Binding site: zinc, catalytic (Cys, His, His, His) (inhibited) #status  
F;122-139,188-227,311-383,390-451,403-422,412-486,463-507,512-562,532-555/Disulfide bond  
F;261,265,331/Binding site: zinc, catalytic (His) (active) #status predicted  
F;262/Active site: Glu #status predicted  
F;297,394/Binding site: carboxylate (Asn) (covalent) #status predicted  
C;Function:  
Query Match 13.1%; Score 99.5; DB 2; Length 582;  
Best Local Similarity 23.4%; Pred. No. 0.35;  
Matches 46; Conservative 19; Mismatches 77; Indels 55; Gaps 7;  
QY 8 AFAIVVSG-----SALAGVYDQVLTTRVVTTHMAHAGSDPS----- 43  
DB 116 AFAQLVMSNWFGLNAPLIAAVEGAYEQWAAADVAAMVGYHSGASAAAEQLVPFQOALQOL 175  
QY 44 -TLSTIYQYGSANAALQSDARKSETTI--TQSGYGN-----GADVQ 83  
DB 176 PNLGIGNIGNANLGGNTGDLNTGNIGNLGSNGRGDANLGSNGNIGSNVGGNVGN 235  
QY 84 ---GADNSTIEL-----TQGNFRNNATIDOWNAKNSDITVQYGGNNAALVNOTASD--- 132  
DB 236 GNFGSGNCRAGLPQSGNVGNLGNLGNLGSNGTNSNVFGNTGNVGTGNAGSGNIGA 295  
QY 133 SSVVVRQVQFGNNATAN 149  
DB 286 GNTGSSNWFGNNGIGN 312  
RESULT 15  
S19916  
leishmanolysin (EC 3.4.24.36) precursor - Leishmania mexicana  
N;Alternate names: surface metalloproteinase glycoprotein gp63  
C;Species: Leishmania mexicana  
C;Date: 10-Sep-1999 #sequence\_revision 10-Sep-1999 #text\_change 10-Sep-1999  
C;Accession: S19916; A48564  
R;Medina-Acosta, E.; Karses, R.E.; Russell, D.  
Submitted to the EMBL Data Library February 1992  
A;Description: Structurally distinct genes for the surface protease (gp63) of Leishmania  
A;Reference number: S19916



GenCore version 5.1.1.6  
Copyright (C) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: August 2, 2004, 14:36:12 ; Search time 5.3 Seconds  
(without alignments)  
1483.508 Million cell updates/sec

Title: US-09-543-407-16

Perfect score: 757

Sequence: 1 MKLLKVAFAAIVVSGSALA.....DSSVNVROVGFNNATANQY 151

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 141681 seqs, 52070155 residues

Total number of hits satisfying chosen parameters: 141681

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt\_42.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	658	86.9	151	1	CSGA_SALTY
2	484	63.9	151	1	CSGA_ECOLI
3	474.5	62.7	152	1	CSGA_ECO57
4	118.5	15.7	151	1	CSGB_ECOLI
5	111.5	14.7	151	1	CSGB_SALTY
6	111.5	14.7	151	1	CSGB_SALTY
7	101.5	13.4	590	1	GP63_LEIDO
8	101	13.3	599	1	GP63_LEICH
9	99.5	13.1	646	1	GP63_LEIME
10	99	13.1	1656	1	OMP8_RICJA
11	97.5	12.9	1034	1	ICEN_PANAN
12	96.5	12.7	602	1	GP63_LEIMA
13	95.5	12.6	1258	1	ICEN_ERWHE
14	94.5	12.5	1322	1	ICEN_PANAN
15	93.5	12.4	1567	1	ICEN_XANCT
16	93	12.3	310	1	HSLO_STRMU
17	91.5	12.1	1655	1	OMP8_RICCN
18	87	11.5	331	1	OM81_NEIMB
19	87	11.5	331	1	OM81_NEIMB
20	85.5	11.3	1210	1	ICEN_PSEFL
21	84.5	11.2	1196	1	ICEV_PSEEX
22	84	11.1	823	1	NSP1_YEAST
23	83.5	11.0	1148	1	ICEK_PSEEX
24	83.5	11.0	1200	1	ICEN_PSEXY
25	82.5	10.9	504	1	NRFL_BRARE
26	82	10.8	955	1	FRU_DROME
27	82	10.8	1025	1	SLAP_CAUCR
28	81.5	10.8	422	1	HEAD_BPAPS
29	81.5	10.8	760	1	YBIL_ECOLI
30	81.5	10.8	1158	1	R114_HUMAN
31	81	10.7	856	1	ALE7_AZOVI
32	80.5	10.6	720	1	G7AC_BREDI
33	80	10.6	503	1	NRFL_HUMAN

#### ALIGNMENTS

##### RESULT 1

ID	CSGA_SALTY	STANDARD;	PRT;	151 AA.
AC	P55225;			
DT	01-OCT-1996 (Rel. 34, Created)			
DT	01-OCT-1996 (Rel. 34, Last sequence update)			
DT	10-OCT-2003 (Rel. 42, Last annotation update)			
DE	Major curlin subunit precursor (Fimbrin SEF17).			
GN	CSGA OR AGFA OR STM1144 OR STY1181 OR T1776.			
OS	Salmonella typhimurium,			
OS	Salmonella typhi, and			
OC	Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;			
OC	Enterobacteriaceae; Salmonella.			
OX	NCBI_TaxID=602, 601, 592;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RC	SPECIES=S.typhimurium; STRAIN=SR-11;			
RX	MEDLINE=98117058; PubMed=9457880;			
RA	Romling U., Bian Z., Hammar M., Sierralta W.D., Normark S.;			
PT	"Curli" fibers are highly conserved to operon structure and regulation.;			
RT	Escherichia coli with respect to operon structure and regulation.;			
RL	J. Bacteriol. 180:722-731(1998).			
RN	[2]			
RP	SEQUENCE FROM N.A.			
RC	SPECIES=S.typhimurium; STRAIN=LT2 / SGSC1412 / ATCC 700720;			
RX	MEDLINE=21534948; PubMed=11677609;			
RA	McClelland M., Sanderson K.E., Spieth J., Clifton S.W., Latreille P.,			
RA	Courtney L., Porwollik S., Ali J., Dante M., Du F., Hou S., Layman D.,			
RA	Leonard S., Nguyen C., Scott K., Holmes A., Grewal N., Mulvaney E.,			
RA	Ryan E., Sun H., Florea L., Miller W., Stoneking T., Nhan M.,			
RA	Waterston R., Wilson R.K.;			
RT	"Complete genome sequence of Salmonella enterica serovar Typhimurium			
RT	LT2.;"			
RL	Nature 413:852-856(2001).			
RN	[3]			
RP	SEQUENCE FROM N.A.			
RC	SPECIES=S.typhi; STRAIN=CT18;			
RX	MEDLINE=21534947; PubMed=11677608;			
RA	Parkhill J., Dougan G., James K.D., Bentley S.D., Holden M.T.G., Sebaihia M.,			
RA	Churcher C., Mungall K.L., Bentley K., Chillingworth T., Connerton P.,			
RA	Baker S., Baeham D., Brooks K., Davies R.M., Dowd L., White N., Farrar J.,			
RA	Cronin A., Davis P., Davies R.M., Dougan G., Holden M.T.G., Sebaihia M.,			
RA	Feltwell T., Hamlin N., Haque A., Hien T.T., Holroyd S., Jagels K.,			
RA	Krogh A., Larsen T.S., Leathem S., Moule S., O'Garra P., Parry C.,			
RA	Quail M.A., Rutherford K., Simmonds M., Skelton J., Stevens K.,			
RA	Whitehead S., Barrell B.G.;			
RT	"Complete genome sequence of a multiple drug resistant Salmonella			
RT	enterica serovar Typhi CT18.;"			
RL	Nature 413:848-852(2001).			
RN	[4]			
RP	SEQUENCE FROM N.A.			
RC	SPECIES=S.typhi; STRAIN=TY2 / ATCC 700931;			
RX	MEDLINE=22531367; PubMed=12644504;			
RA	Deng W., Liou S.-R., Plunkett G. III, Mayhew G.F., Rose D.J.,			
RA	Burland V., Kodyanni V., Schwartz D.C., Blattner F.R.;			

Q9wu00 mus musculus  
Q00689 leishmania  
P07662 pseudomonas  
P14914 rickettsia  
Q53047 x outer mem  
P23818 mus musculus  
P19490 rattus norv  
P75383 mycoplasma  
Q53020 r outer mem  
P30690 neisseria m  
Q59280 corynebacte  
Q9hr50 halobacteri

RT "Comparative genomics of Salmonella enterica serovar Typhi strains Ty2  
 RT and CT18.";  
 J. Bacteriol. 185:2330-2337(2003).  
 [5]

RP SEQUENCE FROM N.A.  
 RC SPECIES-S. enteritidis; STRAIN=27655-3B;  
 RX MEDLINE=96146512; PubMed=8550497;  
 RA Collinson S.K., Clouthier S.C., Doran J.L., Bansen P.A., Kay W.W.;  
 RT "Salmonella enteritidis agfBAC operon encoding thin, aggregative  
 RT fimbriae.";  
 J. Bacteriol. 178:662-667(1996).  
 [6]  
 RP SEQUENCE OF 21-151 FROM N.A.  
 RC SPECIES-S. enteritidis; STRAIN=27655-3B;  
 RX MEDLINE=91310586; PubMed=1677357;  
 RA Collinson S.K., Emeody L., Mueller K.-M., Trust T.J., Kay W.W.;  
 RT "Purification and characterization of thin, aggregative fimbriae from  
 RT Salmonella enteritidis.";  
 J. Bacteriol. 173:4773-4781(1991).  
 CC -1- FUNCTION: CURLIN IS THE STRUCTURAL SUBUNIT OF THE CURLI. CURLI ARE  
 CC COILED SURFACE STRUCTURES THAT ASSEMBLE PREFERENTIALLY AT GROWTH  
 CC TEMPERATURES BELOW 37 DEGREES CELSIUS. CURLI CAN BIND TO  
 CC FIBROBLASTS.  
 CC -1- SIMILARITY: BELONGS TO THE CSGA/CSGB FAMILY.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC -----  
 DR EMBL; AJ002301; CAA05317.1; -;  
 DR EMBL; AE008749; AAL20074.1; -;  
 DR EMBL; AL627269; CAD08268.1; -;  
 DR EMBL; AE016840; AAO69399.1; -;  
 DR EMBL; U43280; AAC43599.1; -;  
 DR PIR; JC6039; JC6039;  
 DR StyGene; SG10608; CSGA.  
 KW Fimbriae; Signal; Complete proteome.  
 FT SIGNAL 1 20  
 FT CHAIN 21 151 MAJOR CURLIN SUBUNIT.  
 FT CONFLICT 134 151 SVMVROVGFNNATANYQ -> DSXTQVQS (IN  
 FT REF. 6).  
 FT SEQUENCE 151 AA; 15305 MW; B7DAC0D16B621359 CRC64;

Query Match 86.9%; Score 658; DB 1; Length 151;  
 Best Local Similarity 90.1%; Pred. No. 4.9e-50;  
 Matches 136; Conservative 3; Mismatches 12; Indels ~ 0; Gaps 0;

QY 1 MKLLKVAAPAAIVVSGSALAGVVDVLTWVTHMAHASGPDSTLSIYQGSANALALQ 60  
 DB 1 MKLLKVAAPAAIVVSGSALAGVVPQWGGGNGGNGSGSPDSTLSIYQGSANALALQ 60  
 QY 61 SDARKSETTITQSGYNGADVGAGDANSTIELTQNGFRNNAIDOWNAKNSDITVQYGG 120  
 DB 61 SDARKSETTITQSGYNGADVGAGDANSTIELTQNGFRNNAIDOWNAKNSDITVQYGG 120  
 QY 121 NNAALVNQATSDSSVMVROVGFNNATANYQ 151  
 DB 121 NNAALVNQATSDSSVMVROVGFNNATANYQ 151

RESULT 2  
 CSGA\_ECOLI STANDARD; PRT; 151 AA.  
 AC P28307;  
 DT 01-DEC-1992 (Rel. 24, Created)  
 DT 01-OCT-1996 (Rel. 34, Last sequence update)  
 DT 28-FEB-2003 (Rel. 41, Last annotation update)  
 DE Major curlin subunit precursor.  
 GN CSQA OR E1042.  
 OS Escherichia coli.  
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;  
 OC Enterobacteriaceae; Escherichia.  
 OX NCBI\_TaxID=562;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=K12 / W3110;  
 RX MEDLINE=93211294; PubMed=8459772;  
 RA Olsen A., Arngqvist A.;  
 RT "The RpoS sigma factor relieves H-NS-mediated transcriptional  
 RT repression of csqA, the subunit gene of fibronectin-binding curli in  
 RT Escherichia coli.";  
 RL Mol. Microbiol. 7:523-536(1993).  
 [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=K12 / MC4100;  
 RX MEDLINE=96414468; PubMed=8817489;  
 RA Hammar M., Arngqvist A., Bian Z., Olsen A., Normark S.;  
 RT "Expression of two csq operons is required for production of  
 RT fibronectin- and congo red-binding curli polymers in Escherichia coli  
 RT K-12.";  
 RL Mol. Microbiol. 18:661-670(1995).  
 [3]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=K12 / MG1655;  
 RX MEDLINE=97426617; PubMed=9278503;  
 RA Blatter F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,  
 RA Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.P.,  
 RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,  
 RA Mau B., Shao Y.;  
 RT "The complete sequence of Escherichia coli K-12.";  
 RL Science 277:1453-1474(1997).  
 [4]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=K12;  
 RX MEDLINE=97061202; PubMed=8905232;  
 RA Oshima T., Aiba H., Baba T., Fujita K., Hayashi K., Honjo A.,  
 RA Ikemoto K., Inada T., Itoh T., Kajihara M., Kanai K., Kashimoto K.,  
 RA Kimura S., Kitagawa M., Makino K., Masuda S., Miki T., Mizobuchi K.,  
 RA Mori H., Motomura K., Nakamura Y., Nishimoto H., Nishio Y., Saito N.,  
 RA Sempel G., Seki Y., Tagami H., Takemoto K., Wada C., Yamamoto Y.,  
 RA Yano M., Horiiuchi T.;  
 RT "A 718-Kb DNA sequence of the Escherichia coli K-12 genome  
 RT corresponding to the 12.7-28.0 min region on the linkage map.";  
 RL DNA Res. 3:137-155(1996).  
 [5]  
 RP SEQUENCE OF 21-40.  
 RC STRAIN=K12 / YVEL;  
 RX MEDLINE=9303873; PubMed=1357528;  
 RA Arngqvist A., Olsen A., Pfeiffer J., Russell D.G., Normark S.;  
 RT "The Crl protein activates cryptic genes for curli formation and  
 RT fibronectin binding in Escherichia coli HB101.";  
 RL Mol. Microbiol. 6:2443-2452(1992).  
 [6]  
 RP SEQUENCE OF 21-31.  
 RX MEDLINE=91310586; PubMed=1677357;  
 RA Collinson S.K., Emeody L., Trust T.J., Kay W.W.;  
 RT "Purification and characterization of thin, aggregative fimbriae from  
 RT Salmonella enteritidis.";  
 RL J. Bacteriol. 173:4773-4781(1991).  
 CC -1- FUNCTION: CURLIN IS THE STRUCTURAL SUBUNIT OF THE CURLI. CURLI ARE  
 CC COILED SURFACE STRUCTURES THAT ASSEMBLE PREFERENTIALLY AT GROWTH  
 CC TEMPERATURES BELOW 37 DEGREES CELSIUS. CURLI CAN BIND TO  
 CC FIBROBLASTS.



CC -1- SIMILARITY: BELONGS TO THE CSGA/CSGB FAMILY.  
CC  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/  
CC or send an email to license@isb-sib.ch).  
CC  
CC EMBL; L04979; AAA23616.1; -  
CC EMBL; X90754; CAA62282.1; -  
CC EMBL; AE000205; AAC74126.1; -  
CC EMBL; D90741; BAA35832.1; -  
CC EMBL; D90742; BAA35840.1; -  
CC PIR; S70788; S70788.  
CC Ecogene; EGI1489; CSGA.  
CC Fimbria; Signal; Complete proteome.  
CC SIGNAL 1 20 MAJOR CURLIN SUBUNIT.  
CC CHAIN 21 151 A -> E (IN REF. 1).  
CC CONFLICT 7 7  
CC SEQUENCE 151 AA; 15049 MW; C003470D208D395F CRC64;  
CC  
CC Query Match 63.9%; Score 484; DB 1; Length 151;  
CC Best Local Similarity 66.9%; Pred. No. 4.8e-35; Indels 0; Gaps 0;  
CC Matches 101; Conservative 18; Mismatches 32;  
CC  
CC QY 1 MKLLKVAFAAIVVSGSALAGVVDQLVTRVTHMAHAGSPDSTLSIYQGSANAALALQ 60  
CC Db 1 MKLLKVAFAAIVVSGSALAGVVDQLVTRVTHMAHAGSPDSTLSIYQGSANAALALQ 60  
CC QY 61 SPARKSETTITQSGYNGADVGADNNTIELTQNGFRNNAIDQWAKNSDITVGYGG 120  
CC Db 61 TDARNSDLTITQGGGNGADVGQSDSSIDLTRQGFNSATLDQWNGKNSMTVRQFGG 120  
CC QY 121 NNAALVNOTASDSSVMVROVGFGNATANOY 151  
CC Db 121 GNGAAVDQTASNSVNVTVQVFGNNATAHQY 151  
CC  
CC RESULT 3  
CC CSGA\_ECO57  
CC ID CSGA\_ECO57 STANDARD; PRT; 152 AA.  
CC AC Q30U4; 2003 (Rel. 41, Created)  
CC DT 28-FEB-2003 (Rel. 41, Last sequence update)  
CC DT 28-FEB-2003 (Rel. 41, Last annotation update)  
CC DE Major curlin subunit precursor.  
CC GN CSGA OR Z1676 OR ECS1420.  
CC OS Escherichia coli O157:H7.  
CC CC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;  
CC CC Enterobacteriaceae; Escherichia.  
CC OX NCBI\_TaxID=83334;  
CC [1]  
CC  
CC SEQUENCE FROM N.A.  
CC STRAIN=O157:H7 / ATCC 43895;  
CC MEDLINE=21218556; PubMed=11319125;  
CC RX Ulrich G.A., Keen J.E., Elder R.O.;  
CC RA "Mutations in the csdG promoter associated with variations in curli  
CC RT expression in certain strains of Escherichia coli O157:H7";  
CC RL Appl. Environ. Microbiol. 67:2367-2370 (2001).  
CC [2]  
CC  
CC SEQUENCE FROM N.A.  
CC STRAIN=O157:H7 / EDL933 / ATCC 700927;  
CC RX MEDLINE=21074935; PubMed=11206551;  
CC RA Perna N.T., Plunkett G. III, Burland V., Mau B., Glasner J.D.,  
CC RA Rose D.J., Mayhew G.F., Evans P.S., Gregor J., Kirkpatrick H.A.,  
CC RA Posfai G., Hackett J., Klink S., Boutin A., Shao Y., Miller L.,  
CC RA Grobeck E.J., Davis N.W., Lim A., Dimalanta E.T., Potamouis K.,  
CC RA Apodaca J., Anantharaman T.S., Lin J., Yen G., Schwartz D.C.,  
CC RA Welch R.A., Blattner F.R.;  
CC RT "Genome sequence of enterohaemorrhagic Escherichia coli O157:H7";  
CC Nature 409:529-533 (2001).

RN SEQUENCE FROM N.A.  
RP STRAIN=O157:H7 / RIMD 050952;  
RC MEDLINE=21156231; PubMed=11258796;  
RX Hayashi T., Makino K., Ohnishi M., Kurokawa K., Ishii K., Yokoyama K.,  
RA Han C.-G., Ohtsubo E., Nakayama K., Murata T., Tanaka M., Tobe T.,  
RA Iida T., Takami H., Honda T., Sasakawa C., Ogasawara N., Yasunaga T.,  
RA Kuhara S., Shiba T., Hattori M., Shingawa H.;  
RA "Complete genome sequence of enterohaemorrhagic Escherichia coli  
RT O157:H7 and genomic comparison with a laboratory strain K-12";  
RL DNA Res. 8:11-22 (2001).  
CC -1- FUNCTION: CURLIN IS THE STRUCTURAL SUBUNIT OF THE CURLI. CURLI ARE  
CC COILED SURFACE STRUCTURES THAT ASSEMBLE PREFERENTIALLY AT GROWTH  
CC TEMPERATURES BELOW 37 DEGREES CELSIUS. CURLI CAN BIND TO  
CC FIBRONECTIN.  
CC -1- SIMILARITY: BELONGS TO THE CSGA/CSGB FAMILY.  
CC  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/  
CC or send an email to license@isb-sib.ch).  
CC  
CC EMBL; AF275733; AAK53212.1; -  
CC EMBL; AE005315; AAC55788.1; -  
CC EMBL; AP002554; BAE34843.1; -  
CC PIR; D90806; D90806.  
CC PIR; H85665; H85665.  
CC Fimbria; Signal; Complete proteome.  
CC SIGNAL 1 20 BY SIMILARITY.  
CC CHAIN 21 152 MAJOR CURLIN SUBUNIT.  
CC SEQUENCE 152 AA; 15099 MW; EE2D2D94DDE91243 CRC64;  
CC  
CC Query Match 62.7%; Score 474.5; DB 1; Length 152;  
CC Best Local Similarity 65.8%; Pred. No. 3.2e-34;  
CC Matches 100; Conservative 19; Mismatches 32; Indels 1; Gaps 1;  
CC  
CC QY 1 MKLLKVAFAAIVVSGSALAGVVDQLVTRVTH-EMAHAGSPDSTLSIYQGSANAALAL 59  
CC Db 1 MKLLKVAFAAIVVSGSALAGVVDQLVTRVTH-EMAHAGSPDSTLSIYQGSANAALAL 60  
CC QY 60 QSDARKSETTITQSGYNGADVGADNNTIELTQNGFRNNAIDQWAKNSDITVGYGG 119  
CC Db 61 QADARNSDLTITQGGGNGADVGQSDSSIDLTRQGFNSATLDQWNGKNSMTVRQFG 120  
CC QY 120 GNGAAVDQTASNSVNVTVQVFGNNATAHQY 151  
CC Db 121 GNGAAVDQTASNSVNVTVQVFGNNATAHQY 152  
CC  
CC RESULT 4  
CC CSGB\_ECOLI  
CC ID CSGB\_ECOLI STANDARD; PRT; 151 AA.  
CC AC P39828;  
CC DT 01-FEB-1995 (Rel. 31, Created)  
CC DT 01-OCT-1996 (Rel. 34, Last sequence update)  
CC DT 28-FEB-2003 (Rel. 41, Last annotation update)  
CC DE Minor curlin subunit precursor.  
CC GN CSGB OR B1041 OR Z1675 OR ECS1419.  
CC OS Escherichia coli, and  
CC OS Escherichia coli O157:H7.  
CC CC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;  
CC CC Enterobacteriaceae; Escherichia.  
CC OX NCBI\_TaxID=562, 83334;  
CC [1]  
CC  
CC SEQUENCE FROM N.A.  
CC STRAIN=K12 / MC4100;  
CC MEDLINE=96414468; PubMed=8817489;  
CC RX Hammar M., Arqvist A., Bian Z., Olsen A., Normark S.;  
CC RA "Expression of two csf operons is required for production of  
CC RT fibronectin- and congo red-binding curli polymers in Escherichia coli

[illegible]

\_\_\_\_\_

DR MEROPS; M08.001; --  
 DR InterPro; IPR006025; Pept M Zn BS.  
 DR InterPro; IPR001577; PeptIdase\_M8.  
 DR Pfam; PF01457; Peptidase\_M8; 1.  
 DR PRINTS; PR00782; LSHMANOLYSIN.  
 DR PROSITE; PS00142; ZINC PROTEASE; 1.  
 KW Hydrolase; Metalloprotease; Glycoprotein; Metal-binding; Zinc;  
 KW Zymogen; Signal; Cell adhesion; GPI-anchor; Lipoprotein.  
 FT SIGNAL 1 39  
 FT PROPEP 40 87  
 FT CHAIN 88 565  
 FT PROPEP 566 590  
 FT METAL 251 251  
 FT ACT SITE 252 252  
 FT METAL 255 255  
 FT METAL 321 321  
 FT DISULFID 112 129  
 FT DISULFID 178 217  
 FT DISULFID 301 373  
 FT DISULFID 380 443  
 FT DISULFID 393 412  
 FT DISULFID 402 477  
 FT DISULFID 454 498  
 FT DISULFID 503 553  
 FT DISULFID 523 546  
 FT CARBOHYD 287 287  
 FT LIPID 565 565  
 SQ SEQUENCE 590 AA; 62950 MW; 0FB315D299659F58 CRC64;  
 Query Match 13.4%; Score 101.5; DB 1; Length 590;  
 Best Local Similarity 28.8%; Pred. No. 0.2;  
 Matches 46; Conservative 20; Mismatches 59; Indels 35; Gaps 9;  
 QY 8 AFAAIVVSGSALAGVYDQVLTVRVVTTHMAHASGPDSTLSIYQYGSANAALQSDAR-- 65  
 DB 226 AVGVINIPAANIAASRYDQVLTVRVVTTHMAHALG---FSVVFPRDARILLESINVRHDF 281  
 QY 66 -----SETTITQS--GYNGA-----DVQGGADNSTIELTQNGFRNATIDQWNAK 109  
 DB 282 DVPVINSSTAVAKAREQYCGTGLEYLEMDQGGAGSAGSHIKM-----RNAQ-DELWAP 334  
 QY 110 NSDITVQYGGNNAALVQNTA---SDSSVMVRQVGFNNA 146  
 DB 335 ASD--AGYSALTMAIFQDLGFFQADFS--KAEEMPWGRNA 371  
 RESULT 8  
 ID GP63 LEICH STANDARD; PRT; 599 AA.  
 AC P15706;  
 DT 01-APR-1990 (Rel. 14, Created)  
 DT 01-APR-1990 (Rel. 14, Last sequence update)  
 DT 10-OCT-2003 (Rel. 42, Last annotation update)  
 DE Leishmanolysin precursor (EC 3.4.24.36) (Cell surface protease)  
 DE (Major surface glycoprotein) (GP63 protein) (Promastigote surface  
 DE endopeptidase).  
 GN GP63  
 OS Leishmania chagasi.  
 OC Eukaryota; Euglenozoa; Kinetoplastida; Trypanosomatidae; Leishmania.  
 OX NCBI\_TaxId=44271;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=90205976; PubMed=2320059;  
 RA Miller R.A., Reed S.G., Parsons M.;  
 RT "Leishmania gp63 molecule implicated in cellular adhesion lacks an  
 RT Arg-Gly-Asp sequence.";  
 RL Mol. Biochem. Parasitol. 39:267-274 (1990).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=92112918; PubMed=1370484;  
 RA Ramamoorthy R., Doneison J.E., Paetz K.E., Maybodi M., Roberts S.C.,  
 RA Wilson M.E.

RT "Three distinct RNAs for the surface protease gp63 are differentially  
 RT expressed during development of Leishmania donovani chagasi  
 RT promastigotes to an infectious form.";  
 RL J. Biol. Chem. 267:1888-1895 (1992).  
 CC -!- FUNCTION: Has an integral role during the infection of macrophages  
 CC in the mammalian host.  
 CC -!- CATALYTIC ACTIVITY: Preference for hydrophobic residues at P1 and  
 CC P1' and basic residues at P2 and P3'. A model nonapeptide is  
 CC cleaved at -Ala-Tyr-|-Leu-Lys-Lys-.  
 CC -!- COFACTOR: Binds 1 zinc ion per subunit (By similarity).  
 CC -!- SUBCELLULAR LOCATION: Attached to the membrane by a GPI-anchor.  
 CC -!- SIMILARITY: Belongs to peptidase family M8.  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (see http://www.isb-sib.ch/announce/  
 CC or send an email to license@isb-sib.ch).  
 CC -----  
 CC EMBL; M80672; AAA29238.1; --  
 CC EMBL; M28527; AAA29235.1; --  
 CC PIR; A44951; A44951.  
 CC HSRP; P08148; 1LML.  
 DR MEROPS; M08.001; --  
 DR InterPro; IPR006025; Pept M Zn BS.  
 DR InterPro; IPR001577; PeptIdase\_M8.  
 DR Pfam; PF01457; Peptidase\_M8; 1.  
 DR PRINTS; PR00782; LSHMANOLYSIN.  
 DR PROSITE; PS00142; ZINC PROTEASE; 1.  
 KW Hydrolase; Metalloprotease; Glycoprotein; Metal-binding; Zinc;  
 KW Zymogen; Signal; Cell adhesion; GPI-anchor; Lipoprotein.  
 FT SIGNAL 1 39  
 FT PROPEP 40 97  
 FT CHAIN 98 574  
 FT PROPEP 575 599  
 FT METAL 261 261  
 FT ACT SITE 262 262  
 FT METAL 265 265  
 FT METAL 331 331  
 FT DISULFID 122 139  
 FT DISULFID 188 227  
 FT DISULFID 311 383  
 FT DISULFID 390 452  
 FT DISULFID 403 422  
 FT DISULFID 412 486  
 FT DISULFID 463 507  
 FT DISULFID 512 562  
 FT DISULFID 532 555  
 FT CARBOHYD 297 297  
 FT CARBOHYD 394 394  
 FT LIPID 574 574  
 SQ SEQUENCE 599 AA; 63848 MW; 746730AB8E2A2B7C CRC64;  
 Query Match 13.3%; Score 101; DB 1; Length 599;  
 Best Local Similarity 30.2%; Pred. No. 0.23;  
 Matches 32; Conservative 12; Mismatches 28; Gaps 3;  
 QY 8 AFAAIVVSGSALAGVYDQVLTVRVVTTHMAHASGPDSTLSIYQYGSANAALQSDAR--- 64  
 DB 236 AVGVINIPAANIAASRYDQVLTVRVVTTHMAHALG-----FSVVFPRDARILLESINVRH 288  
 QY 65 -----KSETITQSGYNGADV-----CGGADNSTIEL 92  
 DB 289 KDFDVPVINSSTAVAKAREQYCGTGLEYLEIEDQGGAGSAGSHIKM 334  
 RESULT 9  
 ID GP63 LEIME  
 ID GP63 LEIME  
 AC P43150;  
 STANDARD; PRT; 646 AA.

DT 01-NOV-1995 (Rel. 32, Created)  
 DT 01-NOV-1995 (Rel. 32, Last sequence update)  
 DT 10-OCT-2003 (Rel. 42, Last annotation update)  
 DE Leishmanolysin C1 precursor (EC 3.4.24.36) (Cell surface protease)  
 DE (Major surface glycoprotein) (GP63 protein) (Promastigote surface endopeptidase).  
 GN GP63-C1.  
 OS Leishmania mexicana.  
 OC Eukaryota; Euclenozoa; Kinetoplastida; Trypanosomatidae; Leishmania.  
 OX NCBI\_TaxID=5685;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=MNYC/BZ/62/M379;  
 RX MEDLINE=93149206; PubMed=8426614;  
 RA Medina-Acosta E., Kress R.E., Russell D.G.;  
 RT "structurally distinct genes for the surface protease of Leishmania mexicana are developmentally regulated."  
 RL Mol. Biochem. Parasitol. 57:31-46(1993).  
 CC -!- FUNCTION: Has an integral role during the infection of macrophages in the mammalian host.  
 CC -!- CATALYTIC ACTIVITY: Preference for hydrophobic residues at P1 and P1', and basic residues at P2 and P3'. A model nonapeptide is cleaved at Ala-Tyr|-Leu-Lys-Lys-.  
 CC -!- COFACTOR: Binds 1 zinc ion per subunit (By similarity).  
 CC -!- DEVELOPMENTAL STAGE: Expressed in both the promastigote and the amastigote forms.  
 CC -!- SIMILARITY: Belongs to peptidase family M8.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 CC EMBL; X64394; CAA45733.1; --  
 DR PIR; S19916; S19916.  
 DR HSP; P08148; ILM.  
 DR MEROPS; M08.001; --  
 DR GlycoSuiteDB; P431150; --  
 DR InterPro; IPR006025; Pept M Zn BS.  
 DR InterPro; IPR001577; Peptidase\_M8.  
 DR Pfam; PF01457; Peptidase\_M8; 1.  
 DR PRINTS; PR00782; LSHMANOLYSIN.  
 DR PROSITE; PS00142; ZINC PROTEASE; 1.  
 DR HydroLase; Metalloprotease; Glycoprotein; Metal-binding; Zinc;  
 KW Zymogen; Signal; Cell adhesion; Multigene family.  
 FT SIGNAL 1 39  
 FT PROPEP 40 102 ACTIVATION PEPTIDE (POTENTIAL).  
 FT CHAIN 103 646 LEISHMANOLYSIN C1.  
 FT METAL 266 266 ZINC (CATALYTIC) (BY SIMILARITY).  
 FT ACT\_SITE 267 267 BY SIMILARITY.  
 FT METAL 270 270 ZINC (CATALYTIC) (BY SIMILARITY).  
 FT METAL 336 336 ZINC (CATALYTIC) (BY SIMILARITY).  
 FT DISULFID 127 144 BY SIMILARITY.  
 FT DISULFID 193 232 BY SIMILARITY.  
 FT DISULFID 316 388 BY SIMILARITY.  
 FT DISULFID 395 458 BY SIMILARITY.  
 FT DISULFID 408 427 BY SIMILARITY.  
 FT DISULFID 417 492 BY SIMILARITY.  
 FT DISULFID 469 513 BY SIMILARITY.  
 FT DISULFID 518 568 BY SIMILARITY.  
 FT DISULFID 538 561 BY SIMILARITY.  
 FT CARBOHYD 86 86 N-LINKED (GLCNAC. .) (POTENTIAL).  
 FT CARBOHYD 297 297 N-LINKED (GLCNAC. .) (POTENTIAL).  
 FT CARBOHYD 399 399 N-LINKED (GLCNAC. .) (POTENTIAL).  
 FT CARBOHYD 409 409 N-LINKED (GLCNAC. .) (POTENTIAL).  
 FT CARBOHYD 433 433 N-LINKED (GLCNAC. .) (POTENTIAL).  
 FT CARBOHYD 445 445 N-LINKED (GLCNAC. .) (POTENTIAL).  
 FT CARBOHYD 466 466 N-LINKED (GLCNAC. .) (POTENTIAL).  
 FT CARBOHYD 501 501 N-LINKED (GLCNAC. .) (POTENTIAL).  
 FT SEQUENCE 646 AA; 69054 MW; FE448DDC78C10B0A CRC64;

Query Match 13.1%; Score 99.5; DB 1; Length 646;  
 Best Local Similarity 30.8%; Pred. No. 0.33;  
 Matches 40; Conservative 15; Mismatches 66; Indels 9; Gaps 4;  
 QY 8 AFRAIVVSGALAGVVDVLTVRVVTTHMAHASGPDSTLSIYQYGSANAALQSDARK-- 65  
 DB 241 AVGVINIPAAINIASRYDQLVTRVVTTHMAHAVGSGTF----FGAVGIQVEVPHLRKDF 296  
 QY 66 SETTITGSGYNGADVGGQGDN-STIELTONGFRNNA--TIDOWNAKNSDITVQYGGNN 122  
 DB 297 NVSVITSTTVAKAREQYGCNSLEYLEIEDQGGAGSAGSHIKMRNAKDELMAPAASGY 356  
 QY 123 AALVYNQTASD 132  
 DB 357 TALTMVAFQD 366

## RESULT 10

OMPB\_RICJA STANDARD; PRT; 1656 AA.  
 ID OMPB\_RICJA  
 AC 006653;  
 DT 30-MAY-2000 (Rel. 39, Created)  
 DT 30-MAY-2000 (Rel. 39, Last sequence update)  
 DT 16-OCT-2001 (Rel. 40, Last annotation update)  
 DE Outer membrane protein B precursor (168 kDa surface-layer protein)  
 DE (Surface protein antigen) (Cell surface antigen 5) (Scas) (rOmpB)  
 DE (rOmp B) [Contains: 120 kDa surface-exposed protein (surface protein antigen) (120 kDa outer membrane protein ompB); 32 kDa beta peptide].  
 GN OMPB.  
 OS Rickettsia japonica.  
 OC Bacteria; Proteobacteria; Alphaproteobacteria; Rickettsiales;  
 OC Rickettsiaceae; Rickettsiae; Rickettsia.  
 OX NCBI\_TaxID=35790;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=YH;  
 RA Uchiyama T.;  
 RT "Sequencing of the gene encoding the protein rOmp B of Rickettsia japonica."  
 RL Submitted (MAY-1997) to the EMBL/GenBank/DBJ databases.  
 CC -!- FUNCTION: THE 120 KDA SURFACE-EXPOSED PROTEIN IS A MAJOR STRUCTURAL PROTEIN WHICH MAY PLAY A ROLE AS A RICKETTSIAL VIRULENCE FACTOR AND/OR IMMUNOGEN DURING INFECTION (BY similarity).  
 CC -!- FUNCTION: THE 32 KDA BETA PEPTIDE MAY SERVE AS A MEMBRANE ANCHOR (BY similarity).  
 CC -!- SUBCELLULAR LOCATION: Cell wall. This bacterium is covered by a S-layer with hexagonal symmetry.  
 CC -!- SIMILARITY: BELONGS TO THE RICKETTSIAE OMPA/OMPB FAMILY.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 CC EMBL; AB003681; BAA20138.1; --  
 DR InterPro; IPR006315; Autotransport.  
 DR InterPro; IPR005546; Autotransporter.  
 DR Pfam; PF03797; Autotransporter\_1.  
 DR TIGRfam; TIGR01414; autotrans\_bar1; 2.  
 KW Antigen; S-layer; Cell wall.  
 FT CHAIN 1 1338 120 kDa SURFACE-EXPOSED PROTEIN.  
 FT DOMAIN 528 533 POLY-GLY.  
 FT SEQUENCE 1656 AA; 168097 MW; 3132A69C9D5D5999F CRC64;

Query Match 13.1%; Score 99; DB 1; Length 1656;  
 Best Local Similarity 24.4%; Pred. No. 1.1;  
 Matches 42; Conservative 23; Mismatches 51; Indels 56; Gaps 8;

QY 6 VAAPAAIVGSGALAGYDOLVTRVTHVTHMAHAGSPDSTLSIYQGSANAALALQSDARK 65  
 Db 509 VLAAGAITLQDSA-----TITGDIQGGG-----GAALQSITLANDATK 547  
 QY 66 SETTITQSG---YGNAGADVGQADNSTIELTQNGFRNNATID-----104  
 Db 548 ---TLTUGGANIISANGTINFANGTILKTST--QNNIVDCDLAIATDQTGVVDASS 602  
 QY 105 QWNAKNSDI--TVGQYQGNNAAL-----VNOTASDSSVMVRQVGFNNAT 147  
 Db 603 LTAQAUTLSGTGIIIGANNITLQGFNIGSKTKLNGNVAINELVIGNGS 654

RESULT 11  
 ICEN PANAN STANDARD; PRT; 1034 AA.  
 AC Q47879;  
 DT 01-NOV-1997 (Rel. 35, Created)  
 DT 01-NOV-1997 (Rel. 35, Last sequence update)  
 DT 16-OCT-2001 (Rel. 40, Last annotation update)  
 DE Ice nucleation protein inau.  
 GN INAU.  
 OS Pantoea ananas (Erwinia uredovora).  
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;  
 OC Enterobacteriaceae; Pantoea.  
 OX NCBI\_TaxID=553;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=KUN-3;  
 RA MEDLINE=94264407; PubMed=7764866;  
 RA Michigami Y., Watabe S., Abe K., Obata H., Azai S.;  
 RT "Cloning and sequencing of an ice nucleation active gene of Erwinia  
 RT uredovora".  
 RL Biosci. Biotechnol. Biochem. 58:762-764(1994).  
 CC -!- FUNCTION: Ice nucleation proteins enable bacteria to nucleate  
 CC crystallization in supercooled water.  
 CC -!- SUBCELLULAR LOCATION: Outer membrane.  
 CC -!- DOMAIN: CONTAINS IMPERFECT REPEATS OF A CONSENSUS OCTAPEPTIDE  
 CC A-G-Y-G-S-T-X-T; FURTHER ON A 16-RESIDUE AND A REGIONAL 48-RESIDUE  
 CC PERIODICITY IS SUPERIMPOSED.  
 CC -!- MISCELLANEOUS: A STRUCTURAL MODEL IS SUGGESTED IN WHICH THE ICE  
 CC NUCLEATION PROTEIN DISPLAYS A SYMMETRY RELATED TO THAT OF ICE.  
 CC -!- SIMILARITY: Belongs to the bacterial ice nucleation protein  
 CC family.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 CC EMBL; D14992; BAA03636.1; --  
 CC PIR; JC2143; JC2143.  
 CC HSP; P06620; IINA.  
 CC InterPro; IPR000258; Ice nucleatn.  
 CC Pfam; PF00818; Ice nucleation; 51.  
 CC PRINTS; PR00327; ICNUCLEATN.  
 CC PROSITE; PS00314; ICE NUCLEATION; 34.  
 KW Ice nucleation; Repeat; Outer membrane.  
 FT DOMAIN 162 993  
 SQ SEQUENCE 1034 AA; 103378 MW; FA222523D3333EADD CRC64;

Query Match 12.9%; Score 97.5; DB 1; Length 1034;  
 Best Local Similarity 28.7%; Pred. No. 0.85;  
 Matches 43; Conservative 24; Mismatches 46; Indels 37; Gaps 11;

QY 12 IVVGSALAGY-DOLVTRVTHVTHMAHAGSPDSTLSIYQGSANAALALQSDARKSETTI 70  
 Db 161 IATYGSTLSGTHQSQLIAGYGSTETA---GDSSTL-IAGYGSTGTA-----GSDSTL 208

QY 71 TQSGYGNADVGQADNSTIELTQNGFRNNATIDQWNAKNSDITVGYG-----GNNAAL 125  
 Db 209 V-AGYGSTQTAGEESSQWA-----GYGSTQT-----CMKGSDLTAG-YGSTGTAGDSSSL 256  
 QY 126 V-----NOTASDSSVMVRQVGFNNATANQ 150  
 Db 257 IAGYGSTQTAGEDSSLT--AGYGSTQTQAQ 284

RESULT 12  
 GP63 LEIMA STANDARD; PRT; 602 AA.  
 AC P08148; P15906;  
 DT 01-AUG-1988 (Rel. 08, Created)  
 DT 01-APR-1990 (Rel. 14, Last sequence update)  
 DT 10-OCT-2003 (Rel. 42, Last annotation update)  
 DE Leishmanolysin precursor (EC 3.4.24.36) (Cell surface protease)  
 DE (Major surface glycoprotein) (GP63 protein) (Promastigote surface  
 DE endopeptidase).  
 GN GP63.  
 OS Leishmania major.  
 OC Eukaryota; Euglenozoa; Kinetoplastida; Trypanosomatidae; Leishmania.  
 OX NCBI\_TaxID=5664;  
 RN [1]  
 RP SEQUENCE FROM N.A., AND SEQUENCE OF 101-123.  
 RX MEDLINE=89154764; PubMed=3346625;  
 RA Bustin L.L., McMaster W.R.;  
 RT "Molecular cloning of the major surface antigen of leishmania".;  
 RL J. Exp. Med. 167:724-729(1988).  
 RN [2]  
 RP REVISIONS.  
 RA Bustin L.L., McMaster W.R.;  
 RL J. Exp. Med. 171:589-589(1990).  
 RN [3]  
 RP GPI-ANCHOR.  
 RX MEDLINE=91009116; PubMed=2145267;  
 RA Schneider P., Ferguson M.A.J., McConville M.J., Mehler A.,  
 RA Homans S.W., Bordier C.;  
 RT "Structure of the glycosyl-phosphatidylinositol membrane anchor of  
 RT the Leishmania major promastigote surface protease".;  
 RL J. Biol. Chem. 265:16955-16964(1990).  
 RN [4]  
 RP X-RAY CRYSTALLOGRAPHY (2.6 ANGSTROMS).  
 RX MEDLINE=95406217; PubMed=7675788;  
 RA Schlagenhauf E., Etges R., Metcalf P.;  
 RT "Crystallization and preliminary X-ray diffraction studies of  
 RT leishmanolysin, the major surface metalloproteinase from Leishmania  
 RT major".;  
 RL Proteins 22:58-66(1995).  
 RN [5]  
 RP X-RAY CRYSTALLOGRAPHY (1.86 ANGSTROMS).  
 RX MEDLINE=98416698; PubMed=9739094;  
 RA Schlagenhauf E., Etges R., Metcalf P.;  
 RT "The crystal structure of the Leishmania major surface proteinase  
 RT leishmanolysin".;  
 RL Structure 6:1035-1046(1998).  
 CC -!- FUNCTION: Has an integral role during the infection of macrophages  
 CC in the mammalian host.  
 CC -!- CATALYTIC ACTIVITY: Preference for hydrophobic residues at P1 and  
 CC P1', and basic residues at P2 and P3'. A model nonapeptide is  
 CC cleaved at -Ala-Tyr-[Leu-Lys-Lys-  
 CC -!- COFACTOR: Binds 1 zinc ion per subunit.  
 CC -!- SUBCELLULAR LOCATION: Attached to the membrane by a GPI-anchor.  
 CC -!- FULLY SATURATED, UNBRANCHED 1-O-ALKYL CHAIN (MAINLY C24:0) AND A  
 CC MIXTURE OF FULLY SATURATED UNBRANCHED 2-O-ACYL CHAINS (C12:0,  
 CC C14:0, C16:0, AND C18:0).  
 CC -!- SIMILARITY: Belongs to peptidase family M8.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way

CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

EMBL; Y00647; CAA68673.1; --  
DR PIR; PLO221; PLO221.  
DR PDB; 1LML; 17-SEP-97.  
DR MEROPS; M08.001; --  
DR InterPro; IPR006025; Pept\_M\_Zn\_BS.  
DR InterPro; IPR001577; Peptidase\_M8.  
DR Pfam; PF01457; Peptidase\_M8; 1.  
DR PRINTS; PR00782; LSHMANOLYSIN.  
DR PROSITE; PS00142; ZINC PROTEASE; 1.  
KW Hydrolase; Metalloprotease; Glycoprotein; Metal-binding; Zinc;  
KW Synonym; Signal; Cell adhesion; GPI-anchor; 3D-structure; Lipoprotein.  
FT SIGNAL 1 39  
FT PROPEP 40 100  
FT CHAIN 101 577  
FT CHAIN 101 577  
FT METAL 264 264  
FT METAL 264 264  
FT ACT\_SITE 265 265  
FT METAL 268 268  
FT METAL 334 334  
FT DISULFID 125 142  
FT DISULFID 191 230  
FT DISULFID 314 386  
FT DISULFID 393 455  
FT DISULFID 406 425  
FT DISULFID 415 489  
FT DISULFID 466 510  
FT DISULFID 515 565  
FT DISULFID 535 558  
FT CARBOHYD 300 300  
FT CARBOHYD 407 407  
FT LIPID 577 577  
FT STRAND 101 102  
FT STRAND 107 108  
FT STRAND 111 114  
FT HELIX 116 119  
FT TURN 121 122  
FT TURN 128 129  
FT STRAND 131 133  
FT STRAND 139 141  
FT HELIX 144 146  
FT HELIX 150 158  
FT TURN 159 159  
FT HELIX 160 169  
FT TURN 170 171  
FT STRAND 172 174  
FT STRAND 177 178  
FT STRAND 180 181  
FT TURN 189 190  
FT HELIX 191 193  
FT HELIX 198 202  
FT TURN 203 203  
FT STRAND 205 206  
FT STRAND 210 215  
FT TURN 221 222  
FT STRAND 226 232  
FT TURN 234 235  
FT STRAND 238 244  
FT HELIX 247 249  
FT TURN 256 269  
FT TURN 270 271  
FT TURN 274 279  
FT TURN 280 281  
FT STRAND 283 286  
FT HELIX 289 291  
FT STRAND 296 299  
FT TURN 302 312  
FT TURN 313 313  
FT TURN 315 316  
FT STRAND 320 322

N-LINKED (GLCNAC. . .) (POTENTIAL).  
N-LINKED (GLCNAC. . .) (POTENTIAL).  
GPI-anchor amidated asparagine.

FT TURN 328 332  
FT STRAND 335 335  
FT TURN 337 339  
FT TURN 341 342  
FT TURN 343 343  
FT TURN 344 345  
FT STRAND 353 353  
FT HELIX 356 364  
FT TURN 365 366  
FT STRAND 369 370  
FT HELIX 372 374  
FT TURN 380 383  
FT HELIX 386 390  
FT STRAND 394 395  
FT TURN 396 397  
FT STRAND 402 404  
FT STRAND 413 414  
FT TURN 417 418  
FT STRAND 421 425  
FT STRAND 428 429  
FT HELIX 435 437  
FT TURN 443 444  
FT STRAND 445 446  
FT TURN 450 454  
FT STRAND 458 465  
FT TURN 466 467  
FT HELIX 470 472  
FT TURN 475 477  
FT HELIX 478 480  
FT TURN 485 486  
FT STRAND 487 494  
FT STRAND 496 496  
FT STRAND 506 516  
FT TURN 517 520  
FT STRAND 521 525  
FT TURN 527 528  
FT STRAND 533 534  
FT TURN 537 538  
FT STRAND 540 542  
FT HELIX 543 545  
FT TURN 546 546  
FT STRAND 550 550  
FT TURN 552 553  
FT STRAND 555 557  
FT HELIX 561 565  
FT TURN 566 567  
FT HELIX 569 572  
FT TURN 573 573  
SQ SEQUENCE 602 AA; 63953 MW; 982EF3245D87C43E CRC64;

Query Match 12.7%; Score 96.5; DB 1; Length 602;  
Best Local Similarity 59.5%; Pred No 0.56;  
Matches 22; Conservative 3; Mismatches 9; Indels 3; Gaps 1;

QY 8 AFAAIVVSGSALAGVVDLVTRVVTHEMAHA---SGP 41  
Db 239 AVGVINIPAAIASRYDQLVTRVVTHEMAHALGFGSP 275

RESULT 13  
ID ICEN\_ERWHE STANDARD; PRT; 1258 AA.  
AC P16239;  
DT 01-APR-1990 (Rel. 14, Created)  
DT 01-APR-1990 (Rel. 14, Last sequence update)  
DT 16-OCT-2001 (Rel. 40, Last annotation update)  
DE Ice nucleation protein.  
GN ICE.  
OS Erwinia herbicola.  
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;  
OC Enterobacteriaceae; Pantoea.  
OX NCBI\_TaxID=549;

RN SEQUENCE FROM N.A.  
 RP STRAIN=ML;  
 RC MEDLINE=90152370; PubMed=2515997;  
 RX Warren G.J., Corotto L.V.;  
 RA "The consensus sequence of ice nucleation proteins from *Erwinia*  
 RT *herbicola*, *Pseudomonas fluorescens* and *Pseudomonas syringae*."  
 RL Gene 85:239-242(1989).  
 CC -!- FUNCTION: Ice nucleation proteins enable bacteria to nucleate  
 CC crystallization in supercooled water.  
 CC -!- SUBCELLULAR LOCATION: Outer membrane.  
 CC -!- DOMAIN: CONTAINS 126 IMPERFECT REPEATS OF A CONSENSUS OCTAPEPTIDE  
 CC A-G-Y-G-S-T-X-T; FURTHER ON A 16-RESIDUE AND A REGIONAL 48-RESIDUE  
 CC PERIODICITY IS SUPERIMPOSED.  
 CC -!- MISCELLANEOUS: A STRUCTURAL MODEL IS SUGGESTED IN WHICH THE ICE  
 CC NUCLEATION PROTEIN DISPLAYS A SYMMETRY RELATED TO THAT OF ICE.  
 CC -!- SIMILARITY: Belongs to the bacterial ice nucleation protein  
 CC family.  
 CC  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC  
 CC EMBL; M26382; AAA24823.1; -  
 CC PIR; JQ0188; JQ0188.  
 CC HSP; P06620; IINA.  
 CC InterPro; IPR000258; Ice\_nucleatn.  
 CC Pfam; PF00818; Ice\_nucleatn; 65.  
 CC PRINTS; PR00327; ICENUCLEATN.  
 CC PROSITE; PS00314; ICE\_NUCLEATION; 45.  
 CC Ice nucleation; Repeat; Outer membrane.  
 CC FT DOMAIN 162 1217 OCTAPEPTIDE PERIODICITY.  
 CC SEQUENCE 1258 AA; 125084 MW; 590E8A130077FBD4 CRC64;  
 CC  
 CC Query Match 12.6%; Score 95.5; DB 1; Length 1258;  
 CC Best Local Similarity 28.7%; Pred. No. 1.6;  
 CC Matches 43; Conservative 24; Mismatches 46; Indels 37; Gaps 11;  
 CC  
 CC 12 IVVSGSALAGVY-DOLVTRVVTHEMAHAGSPDSTLSIYQYGSANALALQSDARKSETTI 70  
 CC 161 IATYGTSLGTHOSQIAGYGSTETA---GDSSTL-AGYGSTGTAGA-----DSTL 208  
 CC 71 TQSGYNGADVGGADNSTITLTQNGFRNNATIDOWNAKNSDITVGYGNNAL 125  
 CC 209 V-AGYGSTQTAGESSQMA-----GYGSTQT-----GMKGSDLTAG-YGSTGTAGDSSSL 256  
 CC 126 V-----NOTASDSSVMVRQVGFNNATANQ 150  
 CC 257 IAGYGSTQTAGESSSLT---AGYGSTQTAAQ 284  
 CC  
 CC RESULT 14  
 CC ICEA PANAN STANDARD; PRT; 1322 AA.  
 CC AC F20465;  
 CC DT 01-FEB-1991 (Rel. 17, Created)  
 CC DT 01-FEB-1991 (Rel. 17, Last sequence update)  
 CC DT 16-OCT-2001 (Rel. 40, Last annotation update)  
 CC DE Ice nucleation protein inaA.  
 CC GN INAA.  
 CC OS Pantoea ananas (*Erwinia uredovora*).  
 CC OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;  
 CC OC Enterobacteriaceae; Pantoea.  
 CC OX NCBI\_TaxID=553;  
 CC [1]  
 CC RN SEQUENCE FROM N.A.  
 CC RP MEDLINE=90092494; PubMed=2599095;  
 CC RX Abe K., Watabe S., Emori Y., Watanabe M., Arai S.;  
 CC "An ice nucleation active gene of *Erwinia ananas*. Sequence similarity

RT nucleation activity.";  
 RL FEBS Lett. 258:297-300(1989).  
 CC -!- FUNCTION: Ice nucleation proteins enable bacteria to nucleate  
 CC crystallization in supercooled water.  
 CC -!- SUBCELLULAR LOCATION: Outer membrane (By similarity).  
 CC -!- DOMAIN: CONTAINS MANY IMPERFECT REPEATS OF THE CONSENSUS  
 CC OCTAPEPTIDE A-G-Y-G-S-T-X-T; FURTHER ON A 16-RESIDUE AND A  
 CC REGIONAL 48-RESIDUE PERIODICITY IS SUPERIMPOSED.  
 CC -!- SIMILARITY: Belongs to the bacterial ice nucleation protein  
 CC family.  
 CC  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC  
 CC EMBL; X17316; CAA35194.1; -  
 CC PIR; S07053; S07053.  
 CC HSP; P06620; IINA.  
 CC InterPro; IPR000258; Ice\_nucleatn.  
 CC Pfam; PF00818; Ice\_nucleatn; 69.  
 CC PRINTS; PR00327; ICENUCLEATN.  
 CC PROSITE; PS00314; ICE\_NUCLEATION; 49.  
 CC Ice nucleation; Repeat; Outer membrane.  
 CC FT DOMAIN 162 1281 OCTAPEPTIDE PERIODICITY.  
 CC SEQUENCE 1322 AA; 131094 MW; 89B0EE24AA837039 CRC64;  
 CC  
 CC Query Match 12.5%; Score 94.5; DB 1; Length 1322;  
 CC Best Local Similarity 29.3%; Pred. No. 2.1;  
 CC Matches 41; Conservative 20; Mismatches 46; Indels 33; Gaps 10;  
 CC  
 CC 12 IVVSGSALAGVY-DOLVTRVVTHEMAHAGSPDSTLSIYQYGSANALALQSDARKSETTI 70  
 CC 161 IATYGTSLGTHOSQIAGYGSTETA---GDSSTL-AGYGSTGTAGA-----GSDSTL 208  
 CC 71 TQSGYNGADVGGADNSTITLTQNGFRNNATIDOWNAKNSDITVGYGNNALVNQTA 130  
 CC 209 V-AGYGSTQTAGESSQMA-----GYGSTQT-----GMKGSDLTAG-YGSTGTAGA-----G 251  
 CC 131 SDSSVMVRQVGFNNATANQ 150  
 CC 252 ADSLSI---AGYGSTQTAGE 268  
 CC  
 CC RESULT 15  
 CC ICEN\_XANCT STANDARD; PRT; 1567 AA.  
 CC AC P18127;  
 CC DT 01-NOV-1990 (Rel. 16, Created)  
 CC DT 01-NOV-1990 (Rel. 16, Last sequence update)  
 CC DT 16-OCT-2001 (Rel. 40, Last annotation update)  
 CC DE Ice nucleation protein.  
 CC GN INAX.  
 CC OS Xanthomonas campestris (pv. translucens).  
 CC OC Bacteria; Proteobacteria; Gammaproteobacteria; Xanthomonadales;  
 CC OC Xanthomonadaceae; Xanthomonas.  
 CC OX NCBI\_TaxID=343;  
 CC [1]  
 CC RN SEQUENCE FROM N.A.  
 CC RP STRAIN=X56S;  
 CC MEDLINE=91080859; PubMed=2259339;  
 CC RA Zhao J., Orser C.S.;  
 CC "Conserved repetition in the ice nucleation gene inaX from  
 CC *Xanthomonas campestris* pv. translucens.";  
 CC Mol. Gen. Genet. 223:163-166(1990).  
 CC -!- FUNCTION: Ice nucleation proteins enable bacteria to nucleate  
 CC crystallization in supercooled water.  
 CC -!- SUBCELLULAR LOCATION: Outer membrane (By similarity).  
 CC -!- DOMAIN: CONTAINS 153 IMPERFECT REPEATS OF THE CONSENSUS



Search completed: August 2, 2004, 14:49:29  
Job time : 6.3 secs



GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: August 2, 2004, 14:39:17 ; Search time 29.7 Seconds  
(without alignments)

1604.150 Million cell updates/sec

Title: US-09-543-407-16

Perfect score: 757

Sequence: 1 MKLLKVAFAAIVVSGSALA.....DSSVMVROVGFNNATANQY 151

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1017041 seqs, 315518202 residues

Total number of hits satisfying chosen parameters: 1017041

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

- 1: SPREMEL.25.\*
- 2: sp\_archaea.\*
- 3: sp\_bacteria.\*
- 4: sp\_fungi.\*
- 5: sp\_human.\*
- 6: sp\_invertebrate.\*
- 7: sp\_mammal.\*
- 8: sp\_mhc.\*
- 9: sp\_organelle.\*
- 10: sp\_phase.\*
- 11: sp\_plant.\*
- 12: sp\_rickettsia.\*
- 13: sp\_virus.\*
- 14: sp\_vertebrate.\*
- 15: sp\_unclassified.\*
- 16: sp\_rvrius.\*
- 17: sp\_bacteriopl.\*
- 18: sp\_archaeap.\*

pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	649	85.7	152	033802	O33802 salmonella
2	563.5	74.4	150	07X243	07X243 citrobacter
3	524	69.2	149	07X240	07X240 citrobacter
4	471.5	62.3	152	08CW63	08CW63 escherichia
5	454.5	60.0	150	07X237	07X237 enterobacte
6	340	44.9	76	Q54069	Q54069 salmonella
7	133	17.6	502	083IH4	083IH4 shewanella
8	120	15.9	160	08CW64	08CW64 escherichia
9	118.5	15.7	151	07UCZ1	07UCZ1 shigella fl
10	118.5	15.7	153	083JI6	083JI6 bradyrhizob
11	116	15.3	171	083JI3	083JI3 bradyrhizob
12	115	15.2	139	083IH3	083IH3 shewanella
13	110	14.5	130	083JI4	083JI4 bradyrhizob
14	108.5	14.3	151	07X244	07X244 citrobacter
15	107.5	14.2	154	083JI5	083JI5 bradyrhizob

17	105.5	13.9	91	2	Q9S3J8	Q9S3J8 escherichia
18	102.5	13.5	152	2	Q7X241	Q7X241 citrobacter
19	101.5	13.4	2087	5	Q8MVM7	Q8MVM7 naegleria g
20	101	13.3	599	5	Q25273	Q25273 leishmania
21	99.5	13.1	582	16	P71868	P71868 mycobacteri
22	99.5	13.1	582	16	Q71868	Q71868 mycobacteri
23	99.5	13.1	641	5	Q8MNM20	Q8MNM20 leishmania
24	99.5	13.1	657	5	Q8MNM21	Q8MNM21 leishmania
25	99	13.1	1613	2	Q840U5	Q840U5 rickettsia
26	99	13.1	1618	2	Q9KKB1	Q9KKB1 rickettsia
27	98.5	13.0	151	2	Q7X238	Q7X238 enterobacte
28	98	12.9	145	16	Q8U6N9	Q8U6N9 agrobacteri
29	98	12.9	157	16	Q88HG0	Q88HG0 pseudomonas
30	97	12.8	1612	2	Q840U6	Q840U6 rickettsia
31	97	12.8	1617	2	Q7X5N9	Q7X5N9 rickettsia
32	96.5	12.7	348	13	Q93397	Q93397 cyprinus ca
33	96.5	12.7	644	5	Q43994	Q43994 leishmania
34	96.5	12.7	1306	2	Q93N36	Q93N36 pantoea ana
35	96.5	12.7	1341	16	Q8ED31	Q8ED31 shewanella
36	96.5	12.7	2016	5	Q9BIT0	Q9BIT0 electreureys
37	96	12.7	598	5	Q25275	Q25275 leishmania
38	96	12.7	641	5	Q8MNM22	Q8MNM22 leishmania
39	96	12.7	641	5	Q8MNM48	Q8MNM48 leishmania
40	96	12.7	641	5	Q8MNM23	Q8MNM23 leishmania
41	96	12.7	641	5	Q8MNM24	Q8MNM24 leishmania
42	96	12.7	641	5	Q8MNM25	Q8MNM25 leishmania
43	95.5	12.6	179	2	Q33801	Q33801 salmonella
44	95.5	12.6	617	5	Q94593	Q94593 leishmania
45	95.5	12.6	1613	2	Q9KKB2	Q9KKB2 israeli tic

ALIGNMENTS

RESULT 1

Q33802	PRELIMINARY;	PRT;	152 AA.
ID	O33802		
AC	O33802		
DT	01-JAN-1998 (TREMEL.05, Created)		
DT	01-JAN-1998 (TREMEL.05, Last sequence update)		
DT	01-DEC-2001 (TREMEL.19, Last annotation update)		
DE	AgfA protein (Fragment).		
GN	AGFA.		
OS	Salmonella typhimurium.		
OC	Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;		
OC	Enterobacteriaceae; Salmonella.		
OX	NCBI_TaxID=602;		
RN	[1]		
RP	SEQUENCE FROM N.A.		
RX	MEDLINE=98053981; PubMed=9393832;		
RA	Sukupolvi S.S., Lorentz R.G., Gordon J.I., Bian Z., Pfeifer J.D.,		
RA	Normark S.J., Rhen M.,		
RT	"Expression of thin, aggregative fimbriae promotes interaction of		
RT	Salmonella typhimurium SR-11 with mouse small intestinal epithelial		
RT	cells."		
RL	Infect. Immun. 65:5320-5325(1997).		
DR	EMBL; AJ000514; CAA04151.1; -		
FT	NON TER 152		
SQ	SEQUENCE 152 AA; 9DA7DADC2364B006 CRC64;		

Query Match	85.7%;	Score 649;	DB 2;	Length 152;
Best Local Similarity	88.7%;	Pred. No. 1.1e-44;		
Matches 134;	Conservative 4;	Mismatches 13;	Indels 0;	Gaps 0;
QY	1	MKLLKVAFAAIVVSGSALA	GVYDQVTRVTVTHMAHASGPDSTLSIVYGSANAALQ	60
Db	1	MKLLKVAFAAIVVSGSALA	GVYDQVTRVTVTHMAHASGPDSTLSIVYGSANAALQ	60
QY	61	SDARKSETTITQSGYNGADV	GGADNSITBELTQNGFRNNTIOWNAKNSDIIVGYGG	120
Db	61	SDARKSETTITQSGYNGADV	GGADNSITBELTQNGFRNNTIOWNAKNSDIIVGYGG	120
QY	121	NNAALVNTASDSSVMVROV	GFNNATANQY 151	

Db 121 NNAALVNOTASDSSVMVROVGFNNAPANQY 151  
|||||

## RESULT 2

Q7X243 ID Q7X243 PRELIMINARY; PRT; 150 AA.  
AC Q7X243;  
DT 01-OCT-2003 (TrEMBLrel. 25, Created)  
DT 01-OCT-2003 (TrEMBLrel. 25, Last sequence update)  
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)  
DE Curlin-csgA protein.  
GN CSGA.  
OS Citrobacter sp. Fec2.  
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;  
OC Enterobacteriaceae; Citrobacter.  
OX NCBI\_TaxID=213763;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=Fec2;  
RA Zogaj X., Bokranz W., Nintz M., Romling U.;  
RT "Production of Cellulose and Curli Fimbriae by Members of the Family  
RT Enterobacteriaceae Isolated from the Human Gastrointestinal Tract.";  
RL Infect. Immun. 72:4151-4158(2003).  
DR EMBL: AJ515700; CAD56672.1; -;  
SQ SEQUENCE 150 AA; 15016 MW; 1D7141B8D6973DC6 CRC64;

Query Match 74.4%; Score 563.5; DB 2; Length 150;  
Best Local Similarity 74.2%; Pred. No. 8e-38;  
Matches 118; Conservative 11; Mismatches 13; Indels 17; Gaps 2;  
QY 1 MLLKVAAPAAIVVSGSALAGVVDQLVTRVVTTH- - - - -AHASGPDSTLSIYQYGS 52  
DB 1 MLLQVAAAPAAIVVSGSALAGVQ- - - - -WGGGGGGGGSSGPESTLSIYQSGV 51  
QY 53 ANAALQSDARKSETTITQSGYNGADVQGGADNSTIELTQNGFRNNATIDQWNAKNSD 112  
DB 52 NNAALQSDARKSDTTIHQNGFNGADVQGGSDNSTIDLTQNGFKNNATIDQWNGKNSD 111  
QY 113 ITVQYGGNNAALVNOTASDSSVMVROVGFNNATANQY 151  
DB 112 ITVQYGGHNAALVNOTASDSSVLVHQVGFNNATANQY 150  
|||||

## RESULT 3

Q7X240 ID Q7X240 PRELIMINARY; PRT; 149 AA.  
AC Q7X240;  
DT 01-OCT-2003 (TrEMBLrel. 25, Created)  
DT 01-OCT-2003 (TrEMBLrel. 25, Last sequence update)  
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)  
DE Curlin-csgA protein.  
GN CSGA.  
OS Citrobacter freundii.  
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;  
OC Enterobacteriaceae; Citrobacter.  
OX NCBI\_TaxID=546;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=Pec4;  
RA Zogaj X., Bokranz W., Nintz M., Romling U.;  
RT "Production of Cellulose and Curli Fimbriae by Members of the Family  
RT Enterobacteriaceae Isolated from the Human Gastrointestinal Tract.";  
RL Infect. Immun. 72:4151-4158(2003).  
DR EMBL: AJ515701; CAD56675.1; -;  
SQ SEQUENCE 149 AA; 15260 MW; 946DD52017F648FD CRC64;

Query Match 69.2%; Score 524; DB 2; Length 149;  
Best Local Similarity 72.8%; Pred. No. 1.2e-34;  
Matches 110; Conservative 14; Mismatches 25; Indels 2; Gaps 1;  
QY 1 MLLKVAAPAAIVVSGSALAGVVDQLVTRVVTTHMAHAGSPDSTLSIYQGSANALALQ 60  
|||||

Db 1 MLLKVAAPAAIVVSGSALAGVVPQGGN- - -HGGGSNYGPDSSLSIYQGSNNSANALQ 58  
QY 61 SDARKSETTITQSGYNGADVQGGADNSTIELTQNGFRNNATIDQWNAKNSDITVQYQYG 120  
DB 59 SDARKSDVTITQHGRGNAGAVVQGGADSTISLQTFQNSATIDQWNAKNAIDISVTFQFG 118  
QY 121 NNAALVNOTASDSSVMVROVGFNNATANQY 151  
DB 119 NGALVNVQNTASDSNVLIQVGFNNATANQH 149  
|||||

## RESULT 4

Q8CW63 ID Q8CW63 PRELIMINARY; PRT; 152 AA.  
AC Q8CW63;  
DT 01-MAR-2003 (TrEMBLrel. 23, Created)  
DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)  
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)  
DE Major curlin subunit precursor.  
GN CSGA OR C1306.  
OS Escherichia coli O6.  
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;  
OC Enterobacteriaceae; Escherichia.  
OX NCBI\_TaxID=217992;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=O6:H1 / CFT073 / ATCC 700928;  
RX MEDLINE=22398234; PubMed=12471157;  
RA Welch R.A., Burland V., Plunkett G. III, Redford P., Roesch P.,  
RA Raško D., Buckles E.L., Liou S.-R., Boutin A., Rackett J., Stroud D.,  
RA Mayhew G.F., Rose D.J., Zhou S., Schwartz D.C., Perna N.T.,  
RA Mobley H.L.T., Donnenberg M.S., Blattner F.R.;  
RT "Extensive mosaic structure revealed by the complete genome sequence  
RT of uropathogenic Escherichia coli.";  
RL Proc. Natl. Acad. Sci. U.S.A. 99:17020-17024(2002).  
DR EMBL: AE016759; AAN79779.1; -;  
KW Complete proteome.  
SQ SEQUENCE 152 AA; 15064 MW; 3BA57F34C1240B83 CRC64;

Query Match 62.3%; Score 471.5; DB 16; Length 152;  
Best Local Similarity 65.8%; Pred. No. 2e-30;  
Matches 100; Conservative 18; Mismatches 33; Indels 1; Gaps 1;  
QY 1 MLLKVAAPAAIVVSGSALAGVVDQLVTRVVTTH-EMAHASGPDSTLSIYQGSANALAL 59  
DB 1 MLLKVAAPAAIVVSGSALAGVVPQYGGGGNGGGGNSGPNSELNIYQGGNSALAQ 60  
QY 60 QSDARKSETTITQSGYNGADVQGGADNSTIELTQNGFRNNATIDQWNAKNSDITVQYQYG 119  
DB 61 QADARNSDLITIHQGGNGADVQGGSDSDSIDLTQRFNGNSALDQWNGKSDTMTVKQFG 120  
QY 120 GNNALVNOTASDSSVMVROVGFNNATANQY 151  
DB 121 GGNGAAVDQTASNSVNVTVQVGFNNATAHQY 152  
|||||

## RESULT 5

Q7X237 ID Q7X237 PRELIMINARY; PRT; 150 AA.  
AC Q7X237;  
DT 01-OCT-2003 (TrEMBLrel. 25, Created)  
DT 01-OCT-2003 (TrEMBLrel. 25, Last sequence update)  
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)  
DE Curlin-csgA protein.  
GN CSGA.  
OS Enterobacter sakazakii.  
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;  
OC Enterobacteriaceae; Enterobacter.  
OX NCBI\_TaxID=28141;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=Fec39;  
RA Zogaj X., Bokranz W., Nintz M., Romling U.;  
|||||

```

RT "production of Cellulose and Curli Fimbriae by Members of the Family
RT Enterobacteriaceae Isolated from the Human Gastrointestinal Tract.";
RL Infect. Immun. 72:4151-4158(2003).
DR EMBL: AJ515702; CAD56678.1; --
SQ SEQUENCE 150 AA; 15112 MW; 5D8BB2D872DF15F3 CRC64;

Query Match 60.0%; Score 454.5; DB 2; Length 150;
Best Local Similarity 62.1%; Pred. No. 4.6e-29;
Matches 95; Conservative 24; Mismatches 29; Indels 5; Gaps 2;

QY 1 MKLLVAFPAATVSGSALAGVVDLVTRVVTHEMAHA--SGPDSTLSIYQYGSANAALA 58
DB 1 MKPIKVAALAAVSGSAGMAGNINQ---CGWGHGHGHHGYPSTLIYQYGGNSALA 57

QY 59 LOSDARKSETTITQSGYNGADVGQADNSTIELTQNGFRNNATIDQWNAKNSDITVQY 118
DB 58 LQTDARNVNLNLSQGGNGADVGQSDSSINLTQNGFNSATLDQWNSKDSVMNVQY 117

QY 119 GNNALVNQTASDSSVMVQVFGNNTANQY 151
DB 118 GGLGALVDQTASNSTVNVYTGFGNHTAQY 150

RESULT 6
Q54069 PRELIMINARY; PRT; 76 AA.
ID Q54069;
AC Q54069;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE SEF17 fimbria (Fragment).
GN AGPA.
OS Salmonella enteritidis.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Salmonella.
OX NCBI_TaxID=592;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=SE30;
RA Cox J.M., Egleros S., Woolcock J.B.;
RT "virulence of Salmonella enteritidis in chickens correlates with
RT colony morphology and expression of SEF17 fimbriae.";
RL Submitted (APR-1996) to the EMBL/GenBank/DBJ databases.
DR EMBL: U53207; AAA98671.1; --
FT NON_TER 1
FT NON_TER 76
SQ SEQUENCE 76 AA; 7704 MW; 2FD5411241A7BCB1 CRC64;

Query Match 44.9%; Score 340; DB 2; Length 76;
Best Local Similarity 97.1%; Pred. No. 3.1e-20;
Matches 67; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 37 HASGPDSTLSIYQYGSANAALQSDARKSETTITQSGYNGADVGQADNSTIELTQNG 96
DB 8 NSSGPDSTLSIYQYGSANAALQSDARKSETTITQSGYNGADVGQADNSTIELTQNG 67

QY 97 FRNNATIDQ 105
DB 68 FRNNATIDQ 76

RESULT 7
Q8BIH4 PRELIMINARY; PRT; 502 AA.
ID Q8BIH4;
AC Q8BIH4;
DT 01-MAR-2003 (TrEMBLrel. 23, Created)
DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE Conserved hypothetical protein.
GN SO0865.
OS Shewanella oneidensis.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Alteromonadales;
OC Alteromonadaceae; Shewanella.

```

```

OX NCBI_TaxID=70853;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=MR-1;
RX MEDLINE=22297686; PubMed=12368813;
RA Heidelberg J.P., Paulsen I.T., Nelson K.E., Gaidos E.J., Nelson W.C.,
RA Read T.D., Eissen J.A., Seshadri R., Ward N., Methe B., Clayton R.A.,
RA Meyer T., Tsapin A., Scott J., Beaman M., Brinkac L., Daugherty S.,
RA DeBoy R.T., Dodson R.J., Durkin A.S., Haft D.H., Kolonay J.F.,
RA Madupu R., Peterson J.D., Umayam L.A., White O., Wolf A.M.,
RA Vamathevan J., Weidman J., Impraim M., Lee K., Berry K., Lee C.,
RA Mueller J., Khouri H., Gill J., Utterback T.R., McDonald L.A.,
RA Feldblyum T.V., Smith H.O., Venter J.C., Nealon K.H., Fraser C.M.;
RT "Genome sequence of the dissimilatory metal ion-reducing bacterium
RT Shewanella oneidensis.";
RL Nat. Biotechnol. 20:1118-1123(2002).
DR EMBL: AE015532; AAN53941.1; --
DR TIGR: SO0865; --
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 502 AA; 52441 MW; D08CA23D6C46B62D CRC64;

Query Match 17.6%; Score 133; DB 16; Length 502;
Best Local Similarity 29.6%; Pred. No. 0.013;
Matches 37; Conservative 21; Mismatches 59; Indels 8; Gaps 3;

QY 32 THEMAHAGPDSTLSIYQYGSANA---ALALQSDARKSETTITQSGYNGADVGQADNS 88
DB 258 TAYLSWTGGDNTVDITQDSDSNTVGDLSLIADTQGDNDITIKQGDSNGAEFQVWGDSN 317

QY 89 TIELTQNGFRNNATIDQWNAKNSDITVQYGGNNAALVNQTASDSSVMVQVFGN---- 144
DB 318 DVDLQKRGDANFATFGAYGTDN-DFDLSSKGDNNELVAFATGEDNSTEISQEGDANFAY 376

QY 145 NATAN 149
DB 377 DATGN 381

RESULT 8
Q8CW64 PRELIMINARY; PRT; 160 AA.
ID Q8CW64;
AC Q8CW64;
DT 01-MAR-2003 (TrEMBLrel. 23, Created)
DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE Minor curlin subunit precursor.
GN CSGB OR C1305.
OS Escherichia coli O6.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Escherichia.
OX NCBI_TaxID=217992;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=O6.H1 / CFT073 / ATCC 700928;
RX MEDLINE=22388234; PubMed=12471157;
RA Welch R.A., Burland V., Plunkett G. III, Redford P., Rossch P.,
RA Rasko D., Buckles E.L., Liou S.-R., Boutin A., Hackett J., Stroud D.,
RA Mayhew G.F., Rose D.J., Zhou S., Schwartz D.C., Perna N.T.,
RA Mobley H.L.T., Donnenberg M.S., Blattner F.R.;
RT "Extensive mosaic structure revealed by the complete genome sequence
RT of uropathogenic Escherichia coli.";
RL Proc. Natl. Acad. Sci. U.S.A. 99:17020-17024(2002).
DR EMBL: AE016759; AAN79778.1; --
KW Complete proteome.
SQ SEQUENCE 160 AA; 16963 MW; 49F68448D979B986 CRC64;

Query Match 15.9%; Score 120; DB 16; Length 160;
Best Local Similarity 28.7%; Pred. No. 0.036;
Matches 41; Conservative 17; Mismatches 67; Indels 18; Gaps 3;

QY 22 VYDQ-----LVTRVTHEMAHAGPDSTLSIYQYGSANAALQSDARKSET 68
DB 1 MYDQVQGDNMKNKLFMLTILGAPGIAAAAGYDLANSEYNF----AVNELSKSSNOAA 56

```

QY 69 TITQSGYNGADVGGQADNSTIETQNGFRNNATIDQWNAKNSDITVGGYGGNNAALVQ 128  
 Db 57 IIGAGTNNNAQLRQGGSKLLTWAGGSSNRKAKIDQTDGYNL-AYIDQAGSANDASISQ 115  
 QY 129 TASDSSVWVRQVFGNNATANOY 151  
 Db 116 GAYGNTAMIIQKSGNKANITQY 138

RESULT 9  
 Q83R07 PRELIMINARY; PRT; 160 AA.  
 ID Q83R07  
 AC Q83R07  
 DT 01-JUN-2003 (TEMBLrel. 24, Created)  
 DT 01-JUN-2003 (TEMBLrel. 24, Last sequence update)  
 DT 01-JUN-2003 (TEMBLrel. 24, Last annotation update)  
 DE Minor curlin subunit precursor, similar to CsgA.  
 GN CsgB OR SF1035.  
 OS Shigella flexneri.  
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;  
 OC Enterobacteriaceae; Shigella.  
 OX NCBI\_TaxID=623;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=301 / Serotype 2a;  
 RX MEDLINE=22272406; PubMed=12384590;  
 RA Jin Q., Yuan Z., Xu J., Wang Y., Shen Y., Lu W., Wang J., Liu H.,  
 RA Yang J., Yang F., Zhang X., Zhang G., Wu H., Qu D., Dong J.,  
 RA Sun L., Xue Y., Zhao A., Gao Y., Zhu J., Kan B., Ding K., Chen S.,  
 RA Cheng H., Yao Z., He B., Chen R., Ma D., Qiang B., Wen Y., Hou Y.,  
 RA Yu J.;  
 RA "Genome sequence of Shigella flexneri 2a: insights into pathogenicity  
 RT through comparison with genomes of Escherichia coli K12 and O157";  
 RL Nucleic Acids Res. 30:4432-4441 (2002).  
 DR EMBL; AF015131; AAN42658.1; -.  
 KW Complete proteome.  
 SQ SEQUENCE 160 AA; 16919 MW; 50269F5268D2A32F CRC64;

Query Match 15.9%; Score 120; DB 16; Length 160;  
 Best Local Similarity 28.7%; Pred. No. 0.036;  
 Matches 41; Conservative 17; Mismatches 67; Indels 18; Gaps 3;  
 QY 22 VYDQ-----LVTRVTHMAHAGSPDSTLSIYQYGSANAALQSDARKSET 68  
 Db 1 MYDQVDNKNKLLPMLTILGPGIAAAGYDLANSEYNF-----AVNELSKSFNQAA 56  
 QY 69 TITQSGYNGADVGGQADNSTIETQNGFRNNATIDQWNAKNSDITVGGYGGNNAALVQ 128  
 Db 57 IIGAGTNNNAQLRQGGSKLLTWAGGSSNRKAKIDQTDGYNL-AYIDQAGSANDASISQ 115  
 QY 129 TASDSSVWVRQVFGNNATANOY 151  
 Db 116 GAYGNTAMIIQKSGNKANITQY 138

RESULT 10  
 Q7UCZ1 PRELIMINARY; PRT; 151 AA.  
 ID Q7UCZ1  
 AC Q7UCZ1  
 DT 01-OCT-2003 (TEMBLrel. 25, Created)  
 DT 01-OCT-2003 (TEMBLrel. 25, Last sequence update)  
 DT 01-OCT-2003 (TEMBLrel. 25, Last annotation update)  
 DE Minor curlin subunit.  
 GN CsgB OR S1108.  
 OS Shigella flexneri.  
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;  
 OC Enterobacteriaceae; Shigella.  
 OX NCBI\_TaxID=623;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=2457T / ATCC 700930 / Serotype 2a;  
 RX MEDLINE=22590274; PubMed=12704152;

RA Wei J., Goldberg M.B., Burland V., Venkatesan M.M., Deng W.,  
 RA Fournier G., Mayhew G.F., Plunkett G. III, Rose D.J., Darling A.,  
 RA Mau B., Perna N.T., Payne S.M., Runyen-Janecky L.J., Zhou S.,  
 RA Schwartz D.C., Blattner F.R.;  
 RT "Complete genome sequence and comparative genomics of Shigella  
 RT flexneri serotype 2a strain 2457T";  
 RL Infect. Immun. 71:2775-2786 (2003).  
 DR EMBL; AE016981; AAP16542.1; -.  
 SQ SEQUENCE 151 AA; 15868 MW; 5D5D266B964014A0 CRC64;

Query Match 15.7%; Score 118.5; DB 16; Length 151;  
 Best Local Similarity 31.6%; Pred. No. 0.045;  
 Matches 37; Conservative 13; Mismatches 62; Indels 5; Gaps 2;  
 QY 35 MAHAGSPDSTLSIYQYGSANAALQSDARKSETTTTQSGYNGADVGGQADNSTIETQ 94  
 Db 18 IAAAGYDLANSEYNF-----AVNELSKSFNQAAIIGAGTNNNAQLRQGGSKLLAVVQ 73  
 QY 95 NGFRNNATIDQWNAKNSDITVGGYGGNNAALVNOTASDSSVWVRQVFGNNATANOY 151  
 Db 74 EGSNRKAKIDQTDGYNL-AYIDQAGSANDASISQAGYNTAMIIQKSGNKANITQY 129

RESULT 11  
 Q89JI6 PRELIMINARY; PRT; 153 AA.  
 ID Q89JI6  
 AC Q89JI6  
 DT 01-JUN-2003 (TEMBLrel. 24, Created)  
 DT 01-JUN-2003 (TEMBLrel. 24, Last sequence update)  
 DT 01-JUN-2003 (TEMBLrel. 24, Last annotation update)  
 DE CsgB protein.  
 GN CsgB OR BL15297.  
 OS Bradyrhizobium japonicum.  
 OC Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;  
 OC Bradyrhizobiaceae; Bradyrhizobium.  
 OX NCBI\_TaxID=375;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=USDA 110;  
 RX MEDLINE=22484998; PubMed=12597275;  
 RA Kaneko T., Nakamura Y., Sato S., Minamisawa K., Uchiumi T.,  
 RA Sasamoto S., Watanabe A., Iidesawa K., Iriguchi M., Kawashima K.,  
 RA Kohara M., Matsumoto M., Shimpoto S., Tsuruoka H., Wada T., Yamada M.,  
 RA Tabata S.;  
 RA "Complete genomic sequence of nitrogen-fixing symbiotic bacterium  
 RT Bradyrhizobium japonicum USDA110";  
 RL DNA Res. 9:189-197 (2002).  
 DR EMBL; AP005954; BAC50562.1; -.  
 KW Complete proteome.  
 SQ SEQUENCE 153 AA; 15991 MW; 4CE71DEAC375145B CRC64;

Query Match 15.7%; Score 118.5; DB 16; Length 153;  
 Best Local Similarity 27.0%; Pred. No. 0.046;  
 Matches 41; Conservative 32; Mismatches 56; Indels 23; Gaps 6;  
 QY 2 KLIKVA-FAATVVGSAAGYVDQVTRVTHMAHAGSPDSTLSIYQYGSANAAL 59  
 Db 10 RVLAVALLAAGAATQASAGSQRSVT-----NPNVSIETIVQFGNDVQPVFTI 58  
 QY 60 QSDARKSETTTQSGYNGADVGGQADNSTIETQNGFRNNATIDQWNAKNSDITVGGY 119  
 Db 59 EENSRYNIARVIG-----GSGIVDATI--IQNGTRYANVIMGG-TTNAAVGQSG 108  
 QY 120 GNNALVNOTASDSSVWVRQVFGNNATANOY 151  
 Db 109 LSNATDITQIGNSTNALLLQIGDMNSGAVRQF 140  
 RESULT 12  
 Q89JI3 PRELIMINARY; PRT; 171 AA.  
 ID Q89JI3  
 AC Q89JI3  
 DT 01-JUN-2003 (TEMBLrel. 24, Created)

```

DT 01-JUN-2003 (Tremblrel. 24, Last sequence update)
DT 01-JUN-2003 (Tremblrel. 24, Last annotation update)
DE CsgA protein.
GN CsgA OR BL15300.
OS Bradyrhizobium japonicum.
OC Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;
OC Bradyrhizobiaceae; Bradyrhizobium.
OX NCBI_TaxID=375;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=USDA 110;
RX MEDLINE=22484998; PubMed=12597275;
RA Kaneko T., Nakamura Y., Sato S., Minamisawa K., Uchiyama T.,
RA Sasamoto S., Watanabe A., Idesawa K., Iriguchi M., Kawashima K.,
RA Kohara M., Matsumoto M., Shimpoo S., Tsuruoka H., Wada T., Yamada M.,
RA Tabata S.;
RT "Complete genomic sequence of nitrogen-fixing symbiotic bacterium
RT Bradyrhizobium japonicum USDA110."
RL DNA Res. 9:189-197(2002).
DR EMBL; AP005954; BAC50565.1; -.
KW Complete proteome.
SQ SEQUENCE 171 AA; 17448 MW; 995DB08C01498381 CRC64;

Query Match 15.3%; Score 116; DB 16; Length 171;
Best Local Similarity 37.5%; Pred. No. 0.083;
Matches 36; Conservative 10; Mismatches 40; Indels 10; Gaps 3;

QY 55 AALALQSDARKSETTIT-OSGVNGADVGQ-GADNSTIELTQNGFRNATIDQWNAKNSD 112
DB 49 AVLAASAQAQANTSTTVQGLVNGSVTQNGLTNDSSSTTQIGILNGASTWQTSSPS- 107

QY 113 ITVGOYGGNNAALVNQTASDSSVMVRQVGFNNATA 148
DB 108 -----LNNVSTVNGVQVNSATTGQVAFNGNSA 136

RESULT 13
Q8BIH3 PRELIMINARY; PRT; 139 AA.
AC Q8BIH3;
DT 01-MAR-2003 (Tremblrel. 23, Created)
DT 01-MAR-2003 (Tremblrel. 23, Last sequence update)
DT 01-MAR-2003 (Tremblrel. 23, Last annotation update)
DE Minor curlin subunit CsgB, putative.
GN Shewanella oneidensis.
OS Bacteria; Proteobacteria; Gammaproteobacteria; Alteromonadales;
OC Alteromonadaceae; Shewanella.
OX NCBI_TaxID=70863;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=MR-1;
RX MEDLINE=22297686; PubMed=12368813;
RA Heidelberg J.F., Paulsen I.T., Nelson K.E., Gaidos E.J., Nelson W.C.,
RA Read T.D., Eisen J.A., Seshadri R., Ward N., Methe B., Clayton R.A.,
RA Meyer T., Tsapin A., Scott J., Beanan M., Brinkac L., Daugherty S.,
RA DeBoy R.T., Dodson R.J., Durkin A.S., Haft D.H., Kolonay J.F.,
RA Madupu R., Peterson J.D., Umayam L.A., White O., Wolf A.M.,
RA Vamathevan J., Weidman J., Impraim M., Lee K., Berry K., Lee C.,
RA Mueller J., Khouri H., Gill J., Utterback T.R., McDonald L.A.,
RA Feldblyum T.V., Smith H.O., Venter J.C., Neillson K.H., Fraser C.M.;
RT "Genome sequence of the dissimilatory metal ion-reducing bacterium
RT Shewanella oneidensis";
RL Nat. Biotechnol. 20:1118-1123(2002).
DR EMBL; AE015532; AAN53942.1; -.
DR TIGR; SO0866; -.
KW Complete proteome.
SQ SEQUENCE 139 AA; 14811 MW; 41EC1CFA76957920 CRC64;

Query Match 15.28; Score 115; DB 16; Length 139;
Best Local Similarity 27.08; Pred. No. 0.077;
Matches 33; Conservative 23; Mismatches 52; Indels 14; Gaps 3;

DT 01-JUN-2003 (Tremblrel. 24, Last sequence update)
DT 01-JUN-2003 (Tremblrel. 24, Last annotation update)
DE CsgA protein.
GN CsgA OR BL15300.
OS Bradyrhizobium japonicum.
OC Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;
OC Bradyrhizobiaceae; Bradyrhizobium.
OX NCBI_TaxID=375;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=USDA 110;
RX MEDLINE=22484998; PubMed=12597275;
RA Kaneko T., Nakamura Y., Sato S., Minamisawa K., Uchiyama T.,
RA Sasamoto S., Watanabe A., Idesawa K., Iriguchi M., Kawashima K.,
RA Kohara M., Matsumoto M., Shimpoo S., Tsuruoka H., Wada T., Yamada M.,
RA Tabata S.;
RT "Complete genomic sequence of nitrogen-fixing symbiotic bacterium
RT Bradyrhizobium japonicum USDA110."
RL DNA Res. 9:189-197(2002).
DR EMBL; AP005954; BAC50565.1; -.
KW Complete proteome.
SQ SEQUENCE 171 AA; 17448 MW; 995DB08C01498381 CRC64;

Query Match 15.3%; Score 116; DB 16; Length 171;
Best Local Similarity 37.5%; Pred. No. 0.083;
Matches 36; Conservative 10; Mismatches 40; Indels 10; Gaps 3;

QY 55 AALALQSDARKSETTIT-OSGVNGADVGQ-GADNSTIELTQNGFRNATIDQWNAKNSD 112
DB 49 AVLAASAQAQANTSTTVQGLVNGSVTQNGLTNDSSSTTQIGILNGASTWQTSSPS- 107

QY 113 ITVGOYGGNNAALVNQTASDSSVMVRQVGFNNATA 148
DB 108 -----LNNVSTVNGVQVNSATTGQVAFNGNSA 136

RESULT 13
Q8BIH3 PRELIMINARY; PRT; 139 AA.
AC Q8BIH3;
DT 01-MAR-2003 (Tremblrel. 23, Created)
DT 01-MAR-2003 (Tremblrel. 23, Last sequence update)
DT 01-MAR-2003 (Tremblrel. 23, Last annotation update)
DE Minor curlin subunit CsgB, putative.
GN Shewanella oneidensis.
OS Bacteria; Proteobacteria; Gammaproteobacteria; Alteromonadales;
OC Alteromonadaceae; Shewanella.
OX NCBI_TaxID=70863;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=MR-1;
RX MEDLINE=22297686; PubMed=12368813;
RA Heidelberg J.F., Paulsen I.T., Nelson K.E., Gaidos E.J., Nelson W.C.,
RA Read T.D., Eisen J.A., Seshadri R., Ward N., Methe B., Clayton R.A.,
RA Meyer T., Tsapin A., Scott J., Beanan M., Brinkac L., Daugherty S.,
RA DeBoy R.T., Dodson R.J., Durkin A.S., Haft D.H., Kolonay J.F.,
RA Madupu R., Peterson J.D., Umayam L.A., White O., Wolf A.M.,
RA Vamathevan J., Weidman J., Impraim M., Lee K., Berry K., Lee C.,
RA Mueller J., Khouri H., Gill J., Utterback T.R., McDonald L.A.,
RA Feldblyum T.V., Smith H.O., Venter J.C., Neillson K.H., Fraser C.M.;
RT "Genome sequence of the dissimilatory metal ion-reducing bacterium
RT Shewanella oneidensis";
RL Nat. Biotechnol. 20:1118-1123(2002).
DR EMBL; AE015532; AAN53942.1; -.
DR TIGR; SO0866; -.
KW Complete proteome.
SQ SEQUENCE 139 AA; 14811 MW; 41EC1CFA76957920 CRC64;

Query Match 15.28; Score 115; DB 16; Length 139;
Best Local Similarity 27.08; Pred. No. 0.077;
Matches 33; Conservative 23; Mismatches 52; Indels 14; Gaps 3;

QY 30 VVTHEMAHAGSPDSTLSIYQYGSANAALALQSDARKSETTITQSGYNGADVGQGDNST 89
DB 32 ITTQALIERSGRDLIDLVOGTANQGIYQSGSDNS-AVYTOAGNDNISLVTQIGTNE 90
QY 90 IELTQNGFRNATIDQWNAKNSDITVGOYGGNNAALVNQTASDSSVMVRQVGFNNATA 149
DB 91 VQLIQVGAQNKASIT-----QIGNDLNLVQLNLGSG-GNFSIQQIADGAASIT 137
QY 150 QY 151
DB 138 QY 139

RESULT 14
Q89JI4 PRELIMINARY; PRT; 130 AA.
AC Q89JI4;
DT 01-JUN-2003 (Tremblrel. 24, Created)
DT 01-JUN-2003 (Tremblrel. 24, Last sequence update)
DT 01-JUN-2003 (Tremblrel. 24, Last annotation update)
DE B115299 protein.
GN B115299.
OS Bradyrhizobium japonicum.
OC Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;
OC Bradyrhizobiaceae; Bradyrhizobium.
OX NCBI_TaxID=375;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=USDA 110;
RX MEDLINE=22484998; PubMed=12597275;
RA Kaneko T., Nakamura Y., Sato S., Minamisawa K., Uchiyama T.,
RA Sasamoto S., Watanabe A., Idesawa K., Iriguchi M., Kawashima K.,
RA Kohara M., Matsumoto M., Shimpoo S., Tsuruoka H., Wada T., Yamada M.,
RA Tabata S.;
RT "Complete genomic sequence of nitrogen-fixing symbiotic bacterium
RT Bradyrhizobium japonicum USDA110."
RL DNA Res. 9:189-197(2002).
DR EMBL; AP005954; BAC50564.1; -.
KW Complete proteome.
SQ SEQUENCE 130 AA; 12699 MW; ACFB2D66A48D260F CRC64;

Query Match 14.5%; Score 110; DB 16; Length 130;
Best Local Similarity 24.5%; Pred. No. 0.18;
Matches 36; Conservative 29; Mismatches 62; Indels 20; Gaps 4;

QY 4 LKVAFAAIVVGSALAGVYDQVTVVTHEMAHAGSPDSTLSIYQYGSANAALALQSDA 63
DB 1 MRITVAVATAIALSALTVDQA-----AGNSASVLQFQTTNSSFISQGS 45
QY 64 RKSETTITQSGYNGADVGQGDNSTIELTQNGFRNATIDQWNAKNSDITVGOYGGNNA 123
DB 46 TSNATTLQFGATNTATTLQGSLLTVNTAVTG-QGGTTA---TASNTALT-GQVGGSNS 100

RESULT 15
Q7X244 PRELIMINARY; PRT; 151 AA.
AC Q7X244;
DT 01-OCT-2003 (Tremblrel. 25, Created)
DT 01-OCT-2003 (Tremblrel. 25, Last sequence update)
DT 01-OCT-2003 (Tremblrel. 25, Last annotation update)
DE Nucleation component of curlin monomers.
GN CSGB.
OS Citrobacter sp. Fec2.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Citrobacter.
OX NCBI_TaxID=213763;
RN [1]
RP SEQUENCE FROM N.A.

```





GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: August 2, 2004, 14:35:42 ; Search time 44.9 Seconds  
(without alignments)  
950.215 Million cell updates/sec

Title: US-09-543-407-18  
Perfect score: 780  
Sequence: 1 MKLLKVAFAAIVVSGSALA.....DSSVMVRQVGFNNATANQY 151

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 1586107 seqs, 282547505 residues

Total number of hits satisfying chosen parameters: 1586107

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : A\_Geneseq\_29Jan04.\*  
1: Geneseq1980s.\*  
2: Geneseq1990s.\*  
3: Geneseq2000s.\*  
4: Geneseq2001s.\*  
5: Geneseq2002s.\*  
6: Geneseq2003as.\*  
7: Geneseq2003bs.\*  
8: Geneseq2004s.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	780	100.0	151	3	AAB36349 Agfa::PT3
2	691	88.6	151	2	AAR74625 Agfa sequ
3	691	88.6	151	3	AAB36341 Salmonell
4	686	87.9	151	2	AAB23570 Salmonell
5	675	86.5	151	3	AAB36353 Agfa::PT3
6	622	79.7	151	3	AAB36350 Agfa::PT3
7	618	79.2	151	3	AAB36348 Agfa::PT3
8	613	78.6	151	3	AAB36346 Agfa::PT3
9	611	78.3	151	3	AAB36347 Agfa::PT3
10	608	77.9	151	3	AAB36352 Agfa::PT3
11	601	77.1	151	3	AAB36354 Agfa::PT3
12	600	76.9	151	3	AAB36351 Agfa::PT3
13	599	76.8	151	3	AAB36355 Agfa::PT3
14	523	67.1	151	3	AAB36343 Escherich
15	518	66.4	151	7	AAR82651 E. coli C
16	506	64.9	120	2	AAR826761 Agfa sequ
17	506	64.9	120	2	AAR23569 Salmonell
18	445	57.1	142	2	AAR52664 Fimbriat
19	373	47.8	122	2	AAR52663 Fimbriat
20	237	30.4	45	3	AAB36316 Salmonell
21	132	16.9	22	3	AAB36318 Salmonell
22	123	15.8	23	3	AAB36321 Salmonell
23	123	15.8	23	3	AAB36326 Salmonell
24	123	15.8	23	3	AAB36338 Salmonell
25	115	14.7	22	3	AAB36325 Salmonell

26	115	14.7	22	3	AAB36339	Aab36339 Salmonell
27	115	14.7	22	3	AAB36320	Aab36320 Salmonell
28	111	14.2	22	3	AAB36322	Aab36322 Salmonell
29	111	14.2	22	3	AAB36327	Aab36327 Salmonell
30	111	14.2	22	3	AAB36337	Aab36337 Salmonell
31	109	14.0	24	7	ABR82644	AbR82644 E. coli c
32	107.5	13.8	151	3	ABR82644	Aab36344 Escherich
33	106.5	13.7	151	3	AAB36342	Aab36342 Salmonell
34	102	13.1	26	7	ABR82649	AbR82649 E. coli v
35	101	12.9	262	4	ABR82649	AbR82649 E. coli c
36	98	12.6	26	7	ABR82645	AbR82645 E. coli c
37	96	12.3	19	3	AAB36323	Aab36323 Salmonell
38	96	12.3	19	3	AAB36336	Aab36336 Salmonell
39	96	12.3	19	3	AAB36328	Aab36328 Salmonell
40	95	12.2	24	7	ABR82647	AbR82647 E. coli c
41	93.5	12.0	287	4	ABR82647	AbR82647 E. coli c
42	93	11.9	974	6	ABU17075	AbU17075 Protein e
43	92	11.8	23	3	AAB36331	Aab36331 Escherich
44	92	11.8	975	6	ADA33477	AdA33477 Acinetoba
45	91.5	11.7	580	5	AAG66008	Aag66008 F. necrop

ALIGNMENTS

RESULT 1  
AAB36349  
ID AAB36349 standard; protein; 151 AA.  
XX  
AC AAB36349;  
XX  
DT 26-FEB-2001 (first entry)  
XX  
DE Agfa::PT3#4 amino acid sequence SEQ ID NO:18.  
XX  
KW Salmonella; agfa; chromosomal gene replacement; fimbrin; epitope;  
KW vaccine; immune response; immunogen.  
XX  
OS Salmonella enteritidis.  
OS Escherichia coli.  
OS Synthetic.  
XX  
PN WO200060102-A2.  
XX  
PD 12-OCT-2000.  
XX  
PF 05-APR-2000; 2000WO-CA000356.  
XX  
PR 05-APR-1999; 99US-0127888P.  
XX  
(UYVI-) UNIV VICTORIA.  
XX  
PI White AP, Doran JL, Collison SK, Kay WW;  
XX  
DR WPI; 2000-672631/65.  
XX  
DR N-PSDB; AAC64625.  
XX  
PT Recombinant agfa gene having a segment replaced by a foreign DNA sequence  
PT which encodes foreign epitope or antigen, expresses recombinant Agfa  
PT protein useful for eliciting immune response in animal.  
XX  
PS Disclosure; Page 136; 139pp; English.  
XX  
CC The present invention describes a recombinant agfa gene (I) where a  
CC segment of the gene has been replaced by a segment of a foreign DNA  
CC sequence which encodes a foreign epitope or antigen. Also described are:  
CC (1) use of thin aggregative fimbriae (SEF17/TAF) nucleation depended  
CC assembly system of strains of Salmonella, Escherichia coli and  
CC Enterobacteriaceae for the production of fimbriae comprising recombinant  
CC Agfa, CsgA and Agfa-homologue fimbria subunits, respectively; (2)  
CC directing recombination of a recombinant gene into the chromosome of the  
CC homologous species; (3) directing recombination of a recombinant gene  
CC back into the chromosome of the homologous species, replacing the native

CC copy of that gene; and (4) eliciting an immune response in an animal,  
 CC comprising separating an amino acid polymer comprising a recombinant Agfa  
 CC protein containing a replacement segment or segments of foreign amino  
 CC acid sequence or sequences grown on a Salmonella, E. coli or  
 CC Enterobacteriaceae host cell, from the host cell and introducing the  
 CC polymer into the animal in conjunction with a carrier or diluent. (1) is  
 CC useful for the expression of recombinant Agfa protein which is useful for  
 CC eliciting an immune response in an animal. In a fimbrial presentation  
 CC system the heterologous antigens are presented in high numbers (up to  
 CC 500,000 copies/cell), the hybrid fimbrial protein possesses both the  
 CC immunogenicity and adhesion properties relevant for an efficient live  
 CC vaccine, the carrier fimbrial subunit proteins are usually strong  
 CC immunogens, which may be important for directing an immune response  
 CC against the inserted epitope, and hybrid fimbriae are easy and  
 CC inexpensive to purify in large amount. The present sequence is given in  
 CC the exemplification of the present invention

SQ Sequence 151 AA;

Query Match 100.0%; Score 780; DB 3; Length 151;  
 Best Local Similarity 100.0%; Pred. No. 1.7e-67;  
 Matches 151; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MLLKVAFAAIVVSGSALAGVVPQWGGGNGGNSGGPDYDQLVTRVVTHEMAHALQ 60  
 DB 1 MLLKVAFAAIVVSGSALAGVVPQWGGGNGGNSGGPDYDQLVTRVVTHEMAHALQ 60

QY 61 SDARKSETTITQSGYNGADVGQADNSTIELTQNGFRNNATIDOWNAKNSDITVGQYGG 120  
 DB 61 SDARKSETTITQSGYNGADVGQADNSTIELTQNGFRNNATIDOWNAKNSDITVGQYGG 120

QY 121 NNAALVNQTASDSSVMVRQVGFNNATANQY 151  
 DB 121 NNAALVNQTASDSSVMVRQVGFNNATANQY 151

RESULT 2  
 AAR74625  
 ID AAR74625 standard; protein; 151 AA.  
 AC AAR74625;  
 DT 25-MAR-2003 (revised)  
 DT 26-JUN-1995 (first entry)  
 XX Agfa sequence.  
 XX Salmonella; Agfa; vaccine.  
 XX Salmonella.  
 XX WO9425598-A2.  
 PD 10-NOV-1994.  
 XX 26-APR-1994; 94WO-TB000207.  
 XX 26-APR-1993; 93US-00054452.  
 XX (UYVI-) UNIV VICTORIA INNOVATION & DEV CORP.  
 PA (KING/) KING J.  
 PI Kay WW, Clouthier SC, Doran JL;  
 XX WPI; 1994-358275/44.  
 DR N-PSDB; AAQ87467.  
 XX Eliciting an immune response to Salmonella - using attenuated Salmonella  
 PT strains, vector constructs, or compns. contg. fimbrial type proteins.  
 PT Disclosure; Fig 7B; 95pp; English.  
 PS The Salmonella Agfa protein and DNA are used in vaccine and genetic

CC immunization compositions, respectively, to elicit an immune response to  
 CC Salmonella in animals (e.g. food producing animals) and humans. (Updated  
 CC on 25-MAR-2003 to correct PN field.)

SQ Sequence 151 AA;

Query Match 88.6%; Score 691; DB 2; Length 151;  
 Best Local Similarity 90.7%; Pred. No. 6.8e-59;  
 Matches 137; Conservative 3; Mismatches 11; Indels 0; Gaps 0;

QY 1 MLLKVAFAAIVVSGSALAGVVPQWGGGNGGNSGGPDYDQLVTRVVTHEMAHALQ 60  
 DB 1 MLLKVAFAAIVVSGSALAGVVPQWGGGNGGNSGGPDYDQLVTRVVTHEMAHALQ 60

QY 61 SDARKSETTITQSGYNGADVGQADNSTIELTQNGFRNNATIDOWNAKNSDITVGQYGG 120  
 DB 61 SDARKSETTITQSGYNGADVGQADNSTIELTQNGFRNNATIDOWNAKNSDITVGQYGG 120

QY 121 NNAALVNQTASDSSVMVRQVGFNNATANQY 151  
 DB 121 NNAALVNQTASDSSVMVRQVGFNNATANQY 151

RESULT 3  
 AAB36341  
 ID AAB36341 standard; protein; 151 AA.  
 AC AAB36341;  
 DT 26-FEB-2001 (first entry)  
 XX Salmonella enteritidis Agfa amino acid sequence SEQ ID NO:5.  
 DE Salmonella; agfa; chromosomal gene replacement; fimbria; epitope;  
 KW vaccine; immune response; immunogen.  
 XX Salmonella enteritidis.  
 OS WO2000060102-A2.  
 PN 12-OCT-2000.  
 PD 05-APR-2000; 2000WO-CA000356.  
 PF 05-APR-1999; 99US-0127888P.  
 PR (UYVI-) UNIV VICTORIA.  
 PA White AP, Doran JL, Collison SK, Kay WW;  
 PI WPI; 2000-672631/65.  
 DR N-PSDB; AAC64617.  
 XX Recombinant agfa gene having a segment replaced by a foreign DNA sequence  
 PT which encodes foreign epitope or antigen, expresses recombinant Agfa  
 PT protein useful for eliciting immune response in animal.  
 XX Disclosure; Page 135; 139pp; English.

The present invention describes a recombinant agfa gene (1) where a  
 segment of the gene has been replaced by a segment of a foreign DNA  
 sequence which encodes a foreign epitope or antigen. Also described are:  
 (1) use of thin aggregative fimbriae (SEF17/TAF) nucleation depended  
 assembly system of strains of Salmonella, Escherichia coli and  
 Enterobacteriaceae for the production of fimbriae comprising recombinant  
 Agfa, CsgA and Agfa-homologue fimbrial subunits, respectively; (2)  
 directing recombination of a recombinant gene into the chromosome of the  
 homologous species; (3) directing recombination of a recombinant gene  
 back into the chromosome of the homologous species, replacing the native  
 copy of that gene; and (4) eliciting an immune response in an animal,  
 comprising separating an amino acid polymer comprising a recombinant Agfa  
 protein containing a replacement segment or segments of foreign amino  
 acid sequence or sequences grown on a Salmonella, E. coli or

CC Enterobacteriaceae host cell, from the host cell and introducing the  
 CC polymer into the animal in conjunction with a carrier or diluent. (1) is  
 CC useful for the expression of recombinant Agfa protein which is useful for  
 CC eliciting an immune response in an animal. In a fimbrial presentation  
 CC system the heterologous antigens are presented in high numbers (up to  
 CC 500,000 copies/cell), the hybrid fimbria protein possesses both the  
 CC immunogenicity and adhesion properties relevant for an efficient live  
 CC vaccine, the carrier fimbrial subunit proteins are usually strong  
 CC immunogens, which may be important for directing an immune response  
 CC against the inserted epitope, and hybrid fimbriae are easy and  
 CC inexpensive to purify in large amount. The present sequence is given in  
 CC the exemplification of the present invention

XX SQ Sequence 151 AA;

Query Match 88.6%; Score 691; DB 3; Length 151;  
 Best Local Similarity 90.7%; Pred. No. 6.8e-59;  
 Matches 137; Conservative 3; Mismatches 11; Indels 0; Gaps 0;

QY 1 MLLKVAFAAIVVSGSALAGVVPQWGGGNGHNGGNSGPDYDQLVTRVTHMAHALQ 60  
 DB 1 MLLKVAFAAIVVSGSALAGVVPQWGGGNGHNGGNSGPDSTLSIYQYGSANAALQ 60

QY 61 SDARKSETTITQSGYNGADVGQADNSTIELTQNGFRNNATIDQWAKNSDITVQYGG 120  
 DB 61 SDARKSETTITQSGYNGADVGQADNSTIELTQNGFRNNATIDQWAKNSDITVQYGG 120

QY 121 NNAALVNQTSADSSVMVRQVGFNNATANQY 151  
 DB 121 NNAALVNQTSADSSVMVRQVGFNNATANQY 151

RESULT 4  
 AAW23570  
 ID AAW23570 standard; protein; 151 AA.  
 XX AC AAW23570;  
 XX DT 25-MAR-2003 (revised)  
 XX DT 29-SEP-1997 (first entry)  
 XX DE Salmonella enteritidis 27655-3b agfa.  
 XX KW Enteropathogenic bacteria; enterobacteria; S.enteritidis; antibody.  
 XX OS Salmonella enteritidis.  
 XX FH Key Location/Qualifiers  
 FT Misc-difference 123  
 FT /note= "Encoded by GCC"  
 XX PN US5635617-A.  
 XX PD 03-JUN-1997.  
 XX PF 26-APR-1994; 94US-00233788.  
 XX PR 26-APR-1993; 93US-00054452.  
 XX PA (UVVI-) UNIV VICTORIA INNOVATION & DEV CORP.

XX PI Collinson SK, Kay WW, Doran JL;  
 XX WPI; 1997-309986/28.  
 XX N-PSDB; AAT74142.  
 XX Isolated Salmonella gene agfa - used for diagnosis of Salmonella or  
 XX enteropathogenic bacteria of the Enterobacteria family.  
 XX Example 2; Fig 7; 85pp; English.

XX The present sequence represents agfa encoded by the full agfa gene  
 CC derived from Salmonella enteritidis 27655-3b. The nucleic acid can be

CC used to provide diagnostic assays for Salmonella and/or enteropathogenic  
 CC bacteria of the family Enterobacteria. It can also be used to provide  
 CC proteins and antibodies which can be used for assays. The nucleic acid  
 CC sequence can be used to provide probes or primers which can specifically  
 CC hybridise to nucleic acid molecules from greater than 99% of Salmonella  
 CC strains that are pathogenic to warm-blooded animals relative to nucleic  
 CC acid molecules from virtually all other microbial organisms. (Updated on  
 XX 25-MAR-2003 to correct PF field.)

XX SQ Sequence 151 AA;

Query Match 87.9%; Score 686; DB 2; Length 151;  
 Best Local Similarity 90.1%; Pred. No. 2.1e-58;  
 Matches 136; Conservative 3; Mismatches 12; Indels 0; Gaps 0;

QY 1 MLLKVAFAAIVVSGSALAGVVPQWGGGNGHNGGNSGPDYDQLVTRVTHMAHALQ 60  
 DB 1 MLLKVAFAAIVVSGSALAGVVPQWGGGNGHNGGNSGPDSTLSIYQYGSANAALQ 60

QY 61 SDARKSETTITQSGYNGADVGQADNSTIELTQNGFRNNATIDQWAKNSDITVQYGG 120  
 DB 61 SDARKSETTITQSGYNGADVGQADNSTIELTQNGFRNNATIDQWAKNSDITVQYGG 120

QY 121 NNAALVNQTSADSSVMVRQVGFNNATANQY 151  
 DB 121 NNAALVNQTSADSSVMVRQVGFNNATANQY 151

RESULT 5  
 AAB36353  
 ID AAB36353 standard; protein; 151 AA.  
 XX AC AAB36353;  
 XX DT 26-FEB-2001 (first entry)  
 XX DE Agfa::PT3#8 amino acid sequence SEQ ID NO:26.  
 XX KW Salmonella; agfa; chromosomal gene replacement; fimbria; epitope;  
 XX KW vaccine; immune response; immunogen.

XX OS Salmonella enteritidis.  
 XX OS Escherichia coli.  
 XX OS Synthetic.  
 XX PN WO200006102-A2.  
 XX PD 12-OCT-2000.  
 XX PF 05-APR-2000; 2000WO-CA000356.  
 XX PR 05-APR-1999; 99US-0127888P.  
 XX PA (UVVI-) UNIV VICTORIA.  
 XX PI White AP, Doran JL, Collison SK, Kay WW;  
 XX WPI; 2000-672631/65.  
 XX N-PSDB; AAC64629.  
 XX Recombinant agfa gene having a segment replaced by a foreign DNA sequence  
 PT which encodes foreign epitope or antigen, expresses recombinant Agfa  
 PT protein useful for eliciting immune response in animal.

XX PS Disclosure; Page 138; 139pp; English.

XX The present invention describes a recombinant agfa gene (1) where a  
 CC segment of the gene has been replaced by a segment of a foreign DNA  
 CC sequence which encodes a foreign epitope or antigen. Also described are:  
 CC (1) use of thin aggregative fimbriae (SEF7/7AF) nucleation depended  
 CC assembly system of strains of Salmonella, Escherichia coli and  
 CC Enterobacteriaceae for the production of fimbriae comprising recombinant  
 CC Agfa, CsgA and Agfa-homologue fimbria subunits, respectively; (2)

CC directing recombination of a recombinant gene into the chromosome of the  
 CC homologous species; (3) directing recombination of a recombinant gene  
 CC back into the chromosome of the homologous species, replacing the native  
 CC copy of that gene; and (4) eliciting an immune response in an animal,  
 CC comprising separating an amino acid polymer comprising a recombinant Agfa  
 CC protein containing a replacement segment or segments of foreign amino  
 CC acid sequence or sequences grown on a Salmonella, E. coli or  
 CC Enterobacteriaceae host cell, from the host cell and introducing the  
 CC polymer into the animal in conjunction with a carrier or diluent. (1) is  
 CC useful for the expression of recombinant Agfa protein which is useful for  
 CC eliciting an immune response in an animal. In a fimbrial presentation  
 CC system the heterologous antigens are presented in high numbers (up to  
 CC 500,000 copies/cell), the hybrid fimbria protein possesses both the  
 CC immunogenicity and adhesion properties relevant for an efficient live  
 CC vaccine, the carrier fimbrial subunit proteins are usually strong  
 CC immunogens, which may be important for directing an immune response  
 CC against the inserted epitope, and hybrid fimbriae are easy and  
 CC inexpensive to purify in large amount. The present sequence is given in  
 CC the exemplification of the present invention  
 CC  
 XX Sequence 151 AA;

Query Match 86.5%; Score 675; DB 3; Length 151;  
 Best Local Similarity 81.9%; Pred. No. 2.4e-57;  
 Matches 136; Conservative 0; Mismatches 0; Indels 30; Gaps 2;  
 QY 1 MLLKVAFAAIVVSGSALAGVVPQGGGNGHNGGNSGPD-----YDQ 45  
 DB 1 MLLKVAFAAIVVSGSALAGVVPQGGGNGHNGGNSGPDSTLSIYQVGSANAALQ 60  
 QY 46 LVTRVVTHEMAHQSDARKSETTITQSGYNGADYVQGGADNSTIELTQNGFRNNATIDQ 105  
 DB 61 LVTRVVTHEMAHA-----GYGNGADYVQGGADNSTIELTQNGFRNNATIDQ 105  
 QY 106 WNAKNSDITVQGGGNNALVNQTSASSVVMVRQVGFNNATANOY 151  
 DB 106 WNAKNSDITVQGGGNNALVNQTSASSVVMVRQVGFNNATANOY 151

RESULT 6  
 AAB36350  
 ID AAB36350 standard; protein; 151 AA.  
 AC AAB36350;  
 XX 26-FEB-2001 (first entry)  
 XX Agfa::PT3#5 amino acid sequence SEQ ID NO:20.  
 XX Salmonella; agfa; chromosomal gene replacement; fimbria; epitope;  
 XX vaccine; immune response; immunogen.  
 XX Salmonella enteritidis.  
 OS Escherichia coli.  
 OS Synthetic.  
 XX WO2000060102-A2.  
 XX 12-OCT-2000.  
 XX 05-APR-2000; 2000WO-CA000356.  
 XX 05-APR-1999; 99US-0127888P.  
 XX (UYVI-) UNIV VICTORIA.  
 XX White AP, Doran JL, Collison SK, Kay WW;  
 PI WPI; 2000-672631/65.  
 DR N-PSDB; AAC64626.  
 XX  
 PT Recombinant agfa gene having a segment replaced by a foreign DNA sequence  
 PT which encodes foreign epitope or antigen, expresses recombinant Agfa

PT protein useful for eliciting immune response in animal.

XX Disclosure; Page 137; 139pp; English.

CC The present invention describes a recombinant agfa gene (1) where a  
 CC segment of the gene has been replaced by a segment of a foreign DNA  
 CC sequence which encodes a foreign epitope or antigen. Also described are:  
 CC (1) use of thin aggregative fimbriae (SFA17/TAF) nucleation depended  
 CC assembly system of strains of Salmonella, Escherichia coli and  
 CC Enterobacteriaceae for the production of fimbriae comprising recombinant  
 CC Agfa, CsgA and Agfa-homologue fimbria subunits, respectively; (2)  
 CC directing recombination of a recombinant gene into the chromosome of the  
 CC homologous species; (3) directing recombination of a recombinant gene  
 CC back into the chromosome of the homologous species, replacing the native  
 CC copy of that gene; and (4) eliciting an immune response in an animal,  
 CC comprising separating an amino acid polymer comprising a recombinant Agfa  
 CC protein containing a replacement segment or segments of foreign amino  
 CC acid sequence or sequences grown on a Salmonella, E. coli or  
 CC Enterobacteriaceae host cell, from the host cell and introducing the  
 CC polymer into the animal in conjunction with a carrier or diluent. (1) is  
 CC useful for the expression of recombinant Agfa protein which is useful for  
 CC eliciting an immune response in an animal. In a fimbrial presentation  
 CC system the heterologous antigens are presented in high numbers (up to  
 CC 500,000 copies/cell), the hybrid fimbria protein possesses both the  
 CC immunogenicity and adhesion properties relevant for an efficient live  
 CC vaccine, the carrier fimbrial subunit proteins are usually strong  
 CC immunogens, which may be important for directing an immune response  
 CC against the inserted epitope, and hybrid fimbriae are easy and  
 CC inexpensive to purify in large amount. The present sequence is given in  
 CC the exemplification of the present invention  
 CC  
 XX Sequence 151 AA;

Query Match 79.7%; Score 622; DB 3; Length 151;  
 Best Local Similarity 73.6%; Pred. No. 3.3e-52;  
 Matches 128; Conservative 0; Mismatches 0; Indels 46; Gaps 2;  
 QY 1 MLLKVAFAAIVVSGSALAGVVPQGGGNGHNGGNSGPD----- 42  
 DB 1 MLLKVAFAAIVVSGSALAGVVPQGGGNGHNGGNSGPDSTLSIYQVGSANAALQ 60  
 QY 43 -----YDQLVTRVVTHEMAHQSDARKSETTITQSGYNGADYVQGGADNSTIELTQNGF 97  
 DB 61 SDARKYDQLVTRVVTHEMAHA-----GQCADNSTIELTQNGF 97  
 QY 98 RNNATIDOWNAKNSDITVQGGGNNALVNQTSASSVVMVRQVGFNNATANOY 151  
 DB 98 RNNATIDOWNAKNSDITVQGGGNNALVNQTSASSVVMVRQVGFNNATANOY 151

RESULT 7  
 AAB36348  
 ID AAB36348 standard; protein; 151 AA.  
 AC AAB36348;  
 XX 26-FEB-2001 (first entry)  
 XX Agfa::PT3#3 amino acid sequence SEQ ID NO:16.  
 XX Salmonella; agfa; chromosomal gene replacement; fimbria; epitope;  
 XX vaccine; immune response; immunogen.  
 XX Salmonella enteritidis.  
 OS Escherichia coli.  
 OS Synthetic.  
 XX WO2000060102-A2.  
 XX 12-OCT-2000.  
 XX 05-APR-2000; 2000WO-CA000356.  
 XX

```

PR 05-APR-1999; 99US-0127888P.
XX (UYVI-) UNIV VICTORIA.
XX White AP, Doran JL, Collison SK, Kay WW;
XX WPI; 2000-672631/65.
XX N-PSDB; AAC64624.
XX Recombinant agfa gene having a segment replaced by a foreign DNA sequence
PT which encodes foreign epitope or antigen, expresses recombinant Agfa
PT protein useful for eliciting immune response in animal.
XX Disclosure; Page 136; 139pp; English.
XX The present invention describes a recombinant agfa gene (I) where a
CC segment of the gene has been replaced by a segment of a foreign DNA
CC sequence which encodes a foreign epitope or antigen. Also described are:
CC (1) use of thin aggregative fimbriae (SEF17/TAF) nucleation depended
CC assembly system of strains of Salmonella, Escherichia coli and
CC Enterobacteriaceae for the production of fimbriae comprising recombinant
CC Agfa, CsgA and Agfa-homologue fimbrial subunits, respectively; (2)
CC directing recombination of a recombinant gene into the chromosome of the
CC homologous species; (3) directing recombination of a recombinant gene
CC back into the chromosome of the homologous species, replacing the native
CC copy of that gene; and (4) eliciting an immune response in an animal
CC comprising separating an amino acid polymer comprising a recombinant Agfa
CC protein containing a replacement segment or segments of foreign amino
CC acid sequence or sequences grown on a Salmonella, E. coli or
CC Enterobacteriaceae host cell, from the host cell and introducing the
CC polymer into the animal in conjunction with a carrier or diluent. (1) is
CC useful for the expression of recombinant Agfa protein which is useful for
CC eliciting an immune response in an animal. In a fimbrial presentation
CC system the heterologous antigens are presented in high numbers (up to
CC 500,000 copies/cell), the hybrid fimbrial protein possesses both the
CC immunogenicity and adhesion properties relevant for an efficient live
CC vaccine, the carrier fimbrial subunit proteins are usually strong
CC immunogens, which may be important for directing an immune response
CC against the inserted epitope, and hybrid fimbriae are easy and
CC inexpensive to purify in large amount. The present sequence is given in
CC the exemplification of the present invention
XX Sequence 151 AA;
XX
XX Query Match 79.2%; Score 618; DB 3; Length 151;
XX Best Local Similarity 76.6%; Pred. No. 8e-52; Indels 40; Gaps 2;
XX Matches 131; Conservative 0; Mismatches 0;
XX
XX QY 1 MKLLKVAAPAAIVVSGSALAGVVPQWGGGNGHNGGNSGGPDYDQLVTRVVTHEMAH 57
XX DB 1 MKLLKVAAPAAIVVSGSALAGV-----YDQLVTRVVTHEMAHAGS 40
XX
XX QY 58 -----ALQSDARKSETTITQSGYNGADVGQADNSTIETQNGFRNN 100
XX DB 41 PDSTLSIYQGSANALALQSDARKSETTITQSGYNGADVGQADNSTIETQNGFRNN 100
XX
XX QY 101 ATIDQWNAKNSDITVGYGNGNAALVNQTSADSSVMVRQVFGGNATANQY 151
XX DB 101 ATIDQWNAKNSDITVGYGNGNAALVNQTSADSSVMVRQVFGGNATANQY 151
XX
XX RESULT 8
XX AAB36346
XX ID AAB36346 standard; protein; 151 AA.
XX AC
XX XX AAB36346;
XX
XX 26-FEB-2001 (first entry)
XX
XX Agfa: PT3#1 amino acid sequence SEQ ID NO:12.
XX
XX Salmonella; agfa; chromosomal gene replacement; fimbrial; epitope;
XX vaccine; immune response; immunogen.
XX
KW

```

```

XX Salmonella enteritidis.
XX Escherichia coli.
XX Synthetic.
XX WO200060102-A2.
XX 12-OCT-2000.
XX
XX 05-APR-2000; 2000WO-CA000356.
XX
XX 05-APR-1999; 99US-0127888P.
XX (UYVI-) UNIV VICTORIA.
XX
XX White AP, Doran JL, Collison SK, Kay WW;
XX WPI; 2000-672631/65.
XX N-PSDB; AAC64622.
XX Recombinant agfa gene having a segment replaced by a foreign DNA sequence
PT which encodes foreign epitope or antigen, expresses recombinant Agfa
PT protein useful for eliciting immune response in animal.
XX Disclosure; Page 135; 139pp; English.
XX The present invention describes a recombinant agfa gene (I) where a
CC segment of the gene has been replaced by a segment of a foreign DNA
CC sequence which encodes a foreign epitope or antigen. Also described are:
CC (1) use of thin aggregative fimbriae (SEF17/TAF) nucleation depended
CC assembly system of strains of Salmonella, Escherichia coli and
CC Enterobacteriaceae for the production of fimbriae comprising recombinant
CC Agfa, CsgA and Agfa-homologue fimbrial subunits, respectively; (2)
CC directing recombination of a recombinant gene into the chromosome of the
CC homologous species; (3) directing recombination of a recombinant gene
CC back into the chromosome of the homologous species, replacing the native
CC copy of that gene; and (4) eliciting an immune response in an animal
CC comprising separating an amino acid polymer comprising a recombinant Agfa
CC protein containing a replacement segment or segments of foreign amino
CC acid sequence or sequences grown on a Salmonella, E. coli or
CC Enterobacteriaceae host cell, from the host cell and introducing the
CC polymer into the animal in conjunction with a carrier or diluent. (1) is
CC useful for the expression of recombinant Agfa protein which is useful for
CC eliciting an immune response in an animal. In a fimbrial presentation
CC system the heterologous antigens are presented in high numbers (up to
CC 500,000 copies/cell), the hybrid fimbrial protein possesses both the
CC immunogenicity and adhesion properties relevant for an efficient live
CC vaccine, the carrier fimbrial subunit proteins are usually strong
CC immunogens, which may be important for directing an immune response
CC against the inserted epitope, and hybrid fimbriae are easy and
CC inexpensive to purify in large amount. The present sequence is given in
CC the exemplification of the present invention
XX Sequence 151 AA;
XX
XX Query Match 78.6%; Score 613; DB 3; Length 151;
XX Best Local Similarity 80.8%; Pred. No. 2.4e-51;
XX Matches 122; Conservative 7; Mismatches 22; Indels 0; Gaps 0;
XX
XX QY 1 MKLLKVAAPAAIVVSGSALAGVVPQWGGGNGHNGGNSGGPDYDQLVTRVVTHEMAH 60
XX DB 1 MKLLKVAAPAAIVVSGSALAGVVPQWGGGNGHNGGNSGGPDSTLSIYQGSANALALQ 60
XX
XX 61 SDARKSETTITQSGYNGADVGQADNSTIETQNGFRNNATIDQWNAKNSDITVGYGG 120
XX DB 61 SDARKSETTITQSGYNGADVGQADNSTIETQNGFRNNATIDQWNAKNSDITVGYGG 120
XX
XX 121 NNAALVNQTSADSSVMVRQVFGGNATANQY 151
XX DB 121 NNAALVNQTSADSSVMVRQVFGGNATANQY 151
XX
XX RESULT 9

```

AAB36347  
ID AAB36347 standard; protein; 151 AA.  
AC AAB36347;  
XX  
XX  
DT 26-FEB-2001 (first entry)  
XX  
XX  
DE Agfa::PT3#2 amino acid sequence SEQ ID NO:14.  
XX  
XX  
KW Salmonella; agfa; chromosomal gene replacement; fimbrin; epitope;  
KW vaccine; immune response; immunogen.  
XX  
XX  
OS Salmonella enteritidis.  
OS Escherichia coli.  
OS Synthetic.  
XX  
XX  
FN WO200060102-A2.  
XX  
XX  
PD 12-OCT-2000.  
XX  
XX  
PF 05-APR-2000; 2000WO-CA000356.  
XX  
XX  
PR 05-APR-1999; 99US-0127888P.  
XX  
XX  
PA (UYVI-) UNIV VICTORIA.  
XX  
XX  
PI White AP, Doran JL, Collison SK, Kay WW;  
XX  
XX  
DR WPI; 2000-672631/65.  
DR N-PSDB; AAC64623.  
XX  
XX  
PT Recombinant agfa gene having a segment replaced by a foreign DNA sequence  
PT which encodes foreign epitope or antigen, expresses recombinant Agfa  
PT protein useful for eliciting immune response in animal.  
XX  
XX  
PS Disclosure; Page 136; 139pp; English.  
XX  
XX  
CC The present invention describes a recombinant agfa gene (I) where a  
CC segment of the gene has been replaced by a segment of a foreign DNA  
CC sequence which encodes a foreign epitope or antigen. Also described are:  
CC (1) use of thin aggregative fimbriae (SEF17/TAF) nucleation depended  
CC assembly system of strains of Salmonella, Escherichia coli and  
CC Enterobacteriaceae for the production of fimbriae comprising recombinant  
CC Agfa, CsgA and Agfa-homologue fimbrin subunits, respectively; (2)  
CC directing recombination of a recombinant gene into the chromosome of the  
CC homologous species; (3) directing recombination of a recombinant gene  
CC back into the chromosome of the homologous species, replacing the native  
CC copy of that gene; and (4) eliciting an immune response in an animal,  
CC comprising separating an amino acid polymer comprising a recombinant Agfa  
CC protein containing a replacement segment or segments of foreign amino  
CC acid sequence or sequences grown on a Salmonella, E. coli or  
CC Enterobacteriaceae host cell, from the host cell and introducing the  
CC polymer into the animal in conjunction with a carrier or diluent. (I) is  
CC useful for the expression of recombinant Agfa protein which is useful for  
CC eliciting an immune response in an animal. In a fimbrial presentation  
CC system the heterologous antigens are presented in high numbers (up to  
CC 500,000 copies/cell) the hybrid fimbrin protein possesses both the  
CC immunogenicity and adhesion properties relevant for an efficient live  
CC vaccine, the carrier fimbrial subunit proteins are usually strong  
CC immunogens, which may be important for directing an immune response  
CC against the inserted epitope, and hybrid fimbriae are easy and  
CC inexpensive to purify in large amount. The present sequence is given in  
CC the exemplification of the present invention  
XX  
SQ Sequence 151 AA;  
Query Match 78.3%; Score 611; DB 3; Length 151;  
Best Local Similarity 81.5%; Pred. No. 3.8e-51;  
Matches 123; Conservative 6; Mismatches 22; Indels 0; Gaps 0;  
QY 1 MKLLKVAAPAAATVWSGALAGVVPWGCGGNGGNGSGGPDYDOLVTRVVTHEMAHALQ 60  
DB 1 MKLLKVAAPAAATVWSGALAGVVPWGCGGNGGNGSGGPDYDOLVTRVVTHEMAHALQ 60

QY 61 SDARKSETTITQSGYNGADVGQADNSTIELTONGFRNNATIDQWNAKNSDITVGYGG 120  
DB 61 SDARKSETTITQSGYNGADVGQADNSTIELTONGFRNNATIDQWNAKNSDITVGYDQ 120  
QY 121 NNAALVNOTASDSVMVVRQVGFNGNATANQY 151  
DB 121 LVTRVVTHEMAHASVMVVRQVGFNGNATANQY 151

RESULT 10  
AAB36352  
ID AAB36352 standard; protein; 151 AA.  
XX  
XX  
AC AAB36352;  
XX  
XX  
DT 26-FEB-2001 (first entry)  
XX  
XX  
DE Agfa::PT3#7 amino acid sequence SEQ ID NO:24.  
XX  
XX  
KW Salmonella; agfa; chromosomal gene replacement; fimbrin; epitope;  
KW vaccine; immune response; immunogen.  
XX  
XX  
OS Salmonella enteritidis.  
OS Escherichia coli.  
OS Synthetic.  
XX  
XX  
FN WO200060102-A2.  
XX  
XX  
PD 12-OCT-2000.  
XX  
XX  
PF 05-APR-2000; 2000WO-CA000356.  
XX  
XX  
PR 05-APR-1999; 99US-0127888P.  
XX  
XX  
PA (UYVI-) UNIV VICTORIA.  
XX  
XX  
PI White AP, Doran JL, Collison SK, Kay WW;  
XX  
XX  
DR WPI; 2000-672631/65.  
DR N-PSDB; AAC64628.  
XX  
XX  
PT Recombinant agfa gene having a segment replaced by a foreign DNA sequence  
PT which encodes foreign epitope or antigen, expresses recombinant Agfa  
PT protein useful for eliciting immune response in animal.  
XX  
XX  
PS Disclosure; Page 138; 139pp; English.  
XX  
XX  
CC The present invention describes a recombinant agfa gene (I) where a  
CC segment of the gene has been replaced by a segment of a foreign DNA  
CC sequence which encodes a foreign epitope or antigen. Also described are:  
CC (1) use of thin aggregative fimbriae (SEF17/TAF) nucleation depended  
CC assembly system of strains of Salmonella, Escherichia coli and  
CC Enterobacteriaceae for the production of fimbriae comprising recombinant  
CC Agfa, CsgA and Agfa-homologue fimbrin subunits, respectively; (2)  
CC directing recombination of a recombinant gene into the chromosome of the  
CC homologous species; (3) directing recombination of a recombinant gene  
CC back into the chromosome of the homologous species, replacing the native  
CC copy of that gene; and (4) eliciting an immune response in an animal,  
CC comprising separating an amino acid polymer comprising a recombinant Agfa  
CC protein containing a replacement segment or segments of foreign amino  
CC acid sequence or sequences grown on a Salmonella, E. coli or  
CC Enterobacteriaceae host cell, from the host cell and introducing the  
CC polymer into the animal in conjunction with a carrier or diluent. (I) is  
CC useful for the expression of recombinant Agfa protein which is useful for  
CC eliciting an immune response in an animal. In a fimbrial presentation  
CC system the heterologous antigens are presented in high numbers (up to  
CC 500,000 copies/cell) the hybrid fimbrin protein possesses both the  
CC immunogenicity and adhesion properties relevant for an efficient live  
CC vaccine, the carrier fimbrial subunit proteins are usually strong  
CC immunogens, which may be important for directing an immune response  
CC against the inserted epitope, and hybrid fimbriae are easy and  
CC inexpensive to purify in large amount. The present sequence is given in  
CC the exemplification of the present invention  
XX  
SQ Sequence 151 AA;

CC the exemplification of the present invention  
 XX Sequence 151 AA;  
 SQ

Query Match 77.9%; Score 608; DB 3; Length 151;  
 Best Local Similarity 82.1%; Pred. No. 7.4e-51;  
 Matches 124; Conservative 5; Mismatches 22; Indels 0; Gaps 0;

QY 1 MKLLKVAFAAIVVSGSALAGVWPQGGGNGHNGGNSGPDYDQLVTRVVTHEMAHALQ 60  
 DB 1 MKLLKVAFAAIVVSGSALAGVWPQGGGNGHNGGNSGPDYDQLVTRVVTHEMAHALQ 60  
 QY 61 SDARKSETTITQSGYNGADVGQADNSTIELTQNGFRNNATIDQWNAKNSDITVGYGG 120  
 DB 61 SDARKSETTITQSGYNGADVGQADNSTIELTQNGFRNNATIDQWNAKNSDITVGYGG 120  
 QY 121 NNAALVNQTASDSSVMVRQVGFNNATANQY 151  
 DB 121 NNAALVNQTASDSSVMVRQVGFNNATANQY 151

RESULT 11  
 AAB36354  
 ID AAB36354 standard; protein; 151 AA.  
 AC AAB36354;  
 XX  
 DT 26-FEB-2001 (first entry)  
 DE Agfa::PT3#9 amino acid sequence SEQ ID NO:28.  
 KW Salmonella; agfa; chromosomal gene replacement; fimbria; epitope;  
 KW vaccine; immune response; immunogen.  
 XX  
 OS Salmonella enteritidis.  
 OS Escherichia coli.  
 OS Synthetic.  
 XX  
 PN WO2000060102-A2.  
 PD 12-OCT-2000.  
 XX  
 PF 05-APR-2000; 2000WO-CA000356.  
 XX  
 PR 05-APR-1999; 99US-0127888P.  
 XX  
 PA (UYVI-) UNIV VICTORIA.  
 XX  
 PI White AP, Doran JL, Collison SK, Kay WW;  
 DR WPI; 2000-672631/65.  
 DR N-PSDB; AAC64630.  
 XX  
 PT Recombinant agfa gene having a segment replaced by a foreign DNA sequence  
 PT which encodes foreign epitope or antigen, expresses recombinant Agfa  
 PT protein useful for eliciting immune response in animal.  
 XX  
 PS Disclosure; Page 138; 139pp; English.  
 XX  
 CC The present invention describes a recombinant agfa gene (I) where a  
 CC segment of the gene has been replaced by a segment of a foreign DNA  
 CC sequence which encodes a foreign epitope or antigen. Also described are:  
 CC (1) use of thin aggregative fimbriae (SEF17/TAF) nucleation depended  
 CC assembly system of strains of Salmonella, Escherichia coli and  
 CC Enterobacteriaceae for the production of fimbriae comprising recombinant  
 CC Agfa, CsgA and Agfa-homologue fimbria subunits, respectively; (2)  
 CC directing recombination of a recombinant gene into the chromosome of the  
 CC homologous species; (3) directing recombination of a recombinant gene  
 CC back into the chromosome of the homologous species, replacing the native  
 CC copy of that gene; and (4) eliciting an immune response in an animal  
 CC comprising separating an amino acid polymer comprising a recombinant Agfa  
 CC protein containing a replacement segment or segments of foreign amino  
 CC acid sequence or sequences grown on a Salmonella, E. coli or

CC Enterobacteriaceae host cell, from the host cell and introducing the  
 CC polymer into the animal in conjunction with a carrier or diluent. (I) is  
 CC useful for the expression of recombinant Agfa protein which is useful for  
 CC eliciting an immune response in an animal. In a fimbrial presentation  
 CC system the heterologous antigens are presented in high numbers (up to  
 CC 500,000 copies/cell), the hybrid fimbria protein possesses both the  
 CC immunogenicity and adhesion properties relevant for an efficient live  
 CC vaccine, the carrier fimbrial subunit proteins are usually strong  
 CC immunogens, which may be important for directing an immune response  
 CC against the inserted epitope, and hybrid fimbriae are easy and  
 CC inexpensive to purify in large amount. The present sequence is given in  
 CC the exemplification of the present invention  
 XX Sequence 151 AA;  
 SQ

Query Match 77.1%; Score 601; DB 3; Length 151;  
 Best Local Similarity 80.8%; Pred. No. 3.5e-50;  
 Matches 122; Conservative 4; Mismatches 25; Indels 0; Gaps 0;

QY 1 MKLLKVAFAAIVVSGSALAGVWPQGGGNGHNGGNSGPDYDQLVTRVVTHEMAHALQ 60  
 DB 1 MKLLKVAFAAIVVSGSALAGVWPQGGGNGHNGGNSGPDYDQLVTRVVTHEMAHALQ 60  
 QY 61 SDARKSETTITQSGYNGADVGQADNSTIELTQNGFRNNATIDQWNAKNSDITVGYGG 120  
 DB 61 SDARKSETTITQSGYNGADVGQADNSTIELTQNGFRNNATIDQWNAKNSDITVGYGG 120  
 QY 121 NNAALVNQTASDSSVMVRQVGFNNATANQY 151  
 DB 121 NNAALVNQTASDSSVMVRQVGFNNATANQY 151

RESULT 12  
 AAB36351  
 ID AAB36351 standard; protein; 151 AA.  
 AC AAB36351;  
 XX  
 DT 26-FEB-2001 (first entry)  
 DE Agfa::PT3#6 amino acid sequence SEQ ID NO:22.  
 KW Salmonella; agfa; chromosomal gene replacement; fimbria; epitope;  
 KW vaccine; immune response; immunogen.  
 XX  
 OS Salmonella enteritidis.  
 OS Escherichia coli.  
 OS Synthetic.  
 XX  
 PN WO2000060102-A2.  
 PD 12-OCT-2000.  
 XX  
 PF 05-APR-2000; 2000WO-CA000356.  
 XX  
 PR 05-APR-1999; 99US-0127888P.  
 XX  
 PA (UYVI-) UNIV VICTORIA.  
 XX  
 PI White AP, Doran JL, Collison SK, Kay WW;  
 DR WPI; 2000-672631/65.  
 DR N-PSDB; AAC64627.  
 XX  
 PT Recombinant agfa gene having a segment replaced by a foreign DNA sequence  
 PT which encodes foreign epitope or antigen, expresses recombinant Agfa  
 PT protein useful for eliciting immune response in animal.  
 XX  
 PS Disclosure; Page 137; 139pp; English.  
 XX  
 CC The present invention describes a recombinant agfa gene (I) where a  
 CC segment of the gene has been replaced by a segment of a foreign DNA  
 CC sequence which encodes a foreign epitope or antigen. Also described are:

CC (1) use of thin aggregative fimbriae (SEF17/TAP) nucleation depended  
 CC assembly system of strains of Salmonella, Escherichia coli and  
 CC Enterobacteriaceae for the production of fimbriae comprising recombinant  
 CC Agfa, CsgA and Agfa-homologue fimbrial subunits, respectively; (2)  
 CC directing recombination of a recombinant gene into the chromosome of the  
 CC homologous species; (3) directing recombination of a recombinant gene  
 CC back into the chromosome of the homologous species, replacing the native  
 CC copy of that gene; and (4) eliciting an immune response in an animal,  
 CC comprising separating an amino acid polymer comprising a recombinant Agfa  
 CC protein containing a replacement segment or segments of foreign amino  
 CC acid sequence or sequences grown on a Salmonella, E. coli or  
 CC Enterobacteriaceae host cell, from the host cell and introducing the  
 CC polymer into the animal in conjunction with a carrier or diluent. (1) is  
 CC useful for the expression of recombinant Agfa protein which is useful for  
 CC eliciting an immune response in an animal. In a fimbrial presentation  
 CC system the heterologous antigens are presented in high numbers (up to  
 CC 500,000 copies/cell), the hybrid fimbrial protein possesses both the  
 CC immunogenicity and adhesion properties relevant for an efficient live  
 CC vaccine, the carrier fimbrial subunit proteins are usually strong  
 CC immunogens, which may be important for directing an immune response  
 CC against the inserted epitope, and hybrid fimbriae are easy and  
 CC inexpensive to purify in large amount. The present sequence is given in  
 CC the exemplification of the present invention

XX SQ Sequence 151 AA;

Query Match 76.9%; Score 600; DB 3; Length 151;  
 Best Local Similarity 81.5%; Pred. No. 4.4e-50;  
 Matches 123; Conservative 5; Mismatches 23; Indels 0; Gaps 0;  
 Qy 1 M K L K V A A F A I V V S G S A L A G V P Q W G G G N H N G S G P D Y D Q L V T R V T H E M A H A L Q 60  
 D b 1 M K L K V A A F A I V V S G S A L A G V P Q W G G G N H N G S G P D Y D Q L V T R V T H E M A H A L Q 60  
 Qy 61 S D A R K S E T T I T Q S G Y G N G A D V G G A D N S T I E L T Q N G F R N N A T I D O W N A K N S D I T V G Y G G 120  
 D b 61 S D A R K S E T T I T Q S G Y G N G A D V G G A D N S T I E L T Q N G F R N N A T I D O W N A K N S D I T V G Y G G 120  
 Qy 121 N N A A L V N Q T A S D S S V M V R Q V G F G N N A T A N Q Y 151  
 D b 121 N N A A L V N Q T A S D S S V M V R Q V G F G N N A T A N Q Y 151

RESULT 13

AA36355 ID AAB36355 standard; protein; 151 AA.

AA36355; AAB36355;

26-FEB-2001 (first entry)

Agfa::PT3#10 amino acid sequence SEQ ID NO:30.

Salmonella; agfa; chromosomal gene replacement; fimbria; epitope;  
 vaccine; immune response; immunogen.

Salmonella enteritidis.

Escherichia coli.

Synthetic.

WO2000060102-A2.

12-OCT-2000.

05-APR-2000; 2000WO-CA000356.

05-APR-1999; 99US-0127888P.

(UYVI-) UNIV VICTORIA.

White AP, Doran JL, Collison SK, Kay WW;

WPI; 2000-672631/55.

DR N-PSDB; AAC64631.

XX Recombinant agfa gene having a segment replaced by a foreign DNA sequence  
 PT which encodes foreign epitope or antigen, expresses recombinant Agfa  
 PT protein useful for eliciting immune response in animal.

PS Disclosure; Page 139; 139pp; English.

XX The present invention describes a recombinant agfa gene (1) where a  
 CC segment of the gene has been replaced by a segment of a foreign DNA  
 CC sequence which encodes a foreign epitope or antigen. Also described are:  
 CC (1) use of thin aggregative fimbriae (SEF17/TAP) nucleation depended  
 CC assembly system of strains of Salmonella, Escherichia coli and  
 CC Enterobacteriaceae for the production of fimbriae comprising recombinant  
 CC Agfa, CsgA and Agfa-homologue fimbrial subunits, respectively; (2)  
 CC directing recombination of a recombinant gene into the chromosome of the  
 CC homologous species; (3) directing recombination of a recombinant gene  
 CC back into the chromosome of the homologous species, replacing the native  
 CC copy of that gene; and (4) eliciting an immune response in an animal,  
 CC comprising separating an amino acid polymer comprising a recombinant Agfa  
 CC protein containing a replacement segment or segments of foreign amino  
 CC acid sequence or sequences grown on a Salmonella, E. coli or  
 CC Enterobacteriaceae host cell, from the host cell and introducing the  
 CC polymer into the animal in conjunction with a carrier or diluent. (1) is  
 CC useful for the expression of recombinant Agfa protein which is useful for  
 CC eliciting an immune response in an animal. In a fimbrial presentation  
 CC system the heterologous antigens are presented in high numbers (up to  
 CC 500,000 copies/cell), the hybrid fimbrial protein possesses both the  
 CC immunogenicity and adhesion properties relevant for an efficient live  
 CC vaccine, the carrier fimbrial subunit proteins are usually strong  
 CC immunogens, which may be important for directing an immune response  
 CC against the inserted epitope, and hybrid fimbriae are easy and  
 CC inexpensive to purify in large amount. The present sequence is given in  
 CC the exemplification of the present invention

XX Sequence 151 AA;

Query Match 76.8%; Score 599; DB 3; Length 151;  
 Best Local Similarity 81.5%; Pred. No. 5.5e-50;  
 Matches 123; Conservative 5; Mismatches 23; Indels 0; Gaps 0;

Qy 1 M K L K V A A F A I V V S G S A L A G V P Q W G G G N H N G S G P D Y D Q L V T R V T H E M A H A L Q 60  
 D b 1 M K L K V A A F A I V V S G S A L A G V P Q W G G G N H N G S G P D Y D Q L V T R V T H E M A H A L Q 60

Qy 61 S D A R K S E T T I T Q S G Y G N G A D V G G A D N S T I E L T Q N G F R N N A T I D O W N A K N S D I T V G Y G G 120  
 D b 61 S D A R K S E T T I T Q S G Y G N G A D V G G A D N S T I E L T Q N G F R N N A T I D O W N A K N S D I T V G Y G G 120

Qy 121 N N A A L V N Q T A S D S S V M V R Q V G F G N N A T A N Q Y 151

D b 121 N N A A L V N Q T A S D S S V M V R Q V G F G N N A T A N Q Y 151

RESULT 14

AAB36343

ID AAB36343 standard; protein; 151 AA.

XX AAB36343;

26-FEB-2001 (first entry)

Escherichia coli CsgA amino acid sequence SEQ ID NO:7.

Salmonella; agfa; chromosomal gene replacement; fimbria; epitope;  
 vaccine; immune response; immunogen.

Escherichia coli.

WO2000060102-A2.

12-OCT-2000.



PF 05-APR-2000; 2000WO-CA000356.  
XX  
PR 05-APR-1999; 99US-0127888P.  
XX  
PA (UYVI-) UNIV VICTORIA.  
XX  
PI White AP, Doran JL, Collison SK, Kay WM;  
XX  
XX WPI; 2000-672631/65.  
DR N-PSDB; AAC64619.  
XX  
XX Recombinant agfa gene having a segment replaced by a foreign DNA sequence  
PT which encodes foreign epitope or antigen, expresses recombinant Agfa  
PT protein useful for eliciting immune response in animal.  
XX  
XX Disclosure; Page 135; 139pp; English.  
XX  
XX The present invention describes a recombinant agfa gene (I) where a  
CC segment of the gene has been replaced by a segment of a foreign DNA  
CC sequence which encodes a foreign epitope or antigen. Also described are:  
CC (1) use of thin aggregative fimbriae (SEPI7/TAF) nucleation depended  
CC assembly system of strains of Salmonella, Escherichia coli and  
CC Enterobacteriaceae for the production of fimbriae comprising recombinant  
CC Agfa, CsgA and Agfa-homologue fimbrial subunits, respectively; (2)  
CC directing recombination of a recombinant gene into the chromosome of the  
CC homologous species; (3) directing recombination of a recombinant gene  
CC back into the chromosome of the homologous species, replacing the native  
CC copy of that gene; and (4) eliciting an immune response in an animal,  
CC comprising separating an amino acid polymer comprising a recombinant Agfa  
CC protein containing a replacement segment or segments of foreign amino  
CC acid sequence or sequences grown on a Salmonella, E. coli or  
CC Enterobacteriaceae host cell, from the host cell and introducing the  
CC polymer into the animal in conjunction with a carrier or diluent. (I) is  
CC useful for the expression of recombinant Agfa protein which is useful for  
CC eliciting an immune response in an animal. In a fimbrial presentation  
CC system the heterologous antigens are presented in high numbers (up to  
CC 500,000 copies/cell), the hybrid fimbrial protein possesses both the  
CC immunogenicity and adhesion properties relevant for an efficient live  
CC vaccine, the carrier fimbrial subunit proteins are usually strong  
CC immunogens, which may be important for directing an immune response  
CC against the inserted epitope, and hybrid fimbriae are easy and  
CC inexpensive to purify in large amount. The present sequence is given in  
CC the exemplification of the present invention  
XX  
SQ Sequence 151 AA;  
Query Match 67.1%; Score 523; DB 3; Length 151;  
Best Local Similarity 68.9%; Pred. No. 1.2e-42;  
Matches 104; Conservative 20; Mismatches 27; Indels 0; Gaps 0;  
QY 1 MKLLKVAAPAAIVVSGSAGALAGVVPQWGGGNGHGGNSGPDYDQLVTRVVTHEMAHALQ 60  
DB 1 MKLLKVAIAAIVFSGSAGALAGVVPQYGGGNGHGGNSGNSSELNIYQYGGNSALALQ 60  
QY 61 SDARKSETTITQSGYNGADVGQADNSTIELTQGFNNATIDQWNAKNSDITVQYGG 120  
DB 61 TDARNSDLTITQGGGNGADVGQSDSSIDLITQRFNGSATLDQWNGKNSMTVKQFGG 120  
QY 121 NNAALVNQTSADSSVMVQVGFNNATANQY 151  
DB 121 GNGAAVDQTASNSVNVTVQVGFNNATAHQY 151  
RESULT 15  
ABR82651  
ID ABR82651 standard; protein; 151 AA.  
XX  
XX ABR82651;  
XX  
XX 04-DEC-2003 (first entry)  
XX  
XX E. coli CsgA subunit 15 kDa protein.  
XX

KW Plasma protein; immune response; antibacterial; vaccine; gene therapy.  
XX Escherichia coli.  
OS  
PN WO2003064446-A2.  
XX  
XX 07-AUG-2003.  
XX  
XX 30-JAN-2003; 2003WO-EP000943.  
XX  
XX 31-JAN-2002; 2002GB-00002275.  
XX  
XX (HANS-) HANSA MEDICAL RES AB.  
XX  
XX Bjoerck L, Olsen A, Wikstroem M, Herwald H;  
XX  
XX WPI; 2003-646136/61.  
DR N-PSDB; ACF36153.  
DR  
XX New isolated peptide capable of binding a mammalian plasma protein,  
PT useful in the manufacture of a medicament for the prevention and/or  
PT treatment of a bacterial infection, such as Escherichia coli, Salmonella  
PT or Shigella infections.  
XX  
XX Disclosure; Page 41-42; 42pp; English.  
XX  
XX The invention relates to an isolated peptide capable of binding a  
CC mammalian plasma protein or of generating an immune response in a mammal  
CC selected from sequences shown in ABR82642, ABR82648-49. The peptide or  
CC antibody is useful for treating a bacterial infection in a human or  
CC animal or in the manufacture of a medicament for the prophylactic  
CC treatment of a bacterial infection, such as Escherichia coli, Salmonella  
CC or Shigella infection. The peptide that is immobilized on a solid support  
CC is also useful as a reagent for determining the ability of a plasma  
CC protein to bind to bacteria. The present sequence represents an E. coli  
CC 15 kDa protein  
XX  
SQ Sequence 151 AA;  
Query Match 66.4%; Score 518; DB 7; Length 151;  
Best Local Similarity 68.2%; Pred. No. 3.8e-42;  
Matches 103; Conservative 20; Mismatches 28; Indels 0; Gaps 0;  
QY 1 MKLLKVAAPAAIVVSGSAGALAGVVPQWGGGNGHGGNSGPDYDQLVTRVVTHEMAHALQ 60  
DB 1 MKLLKVEAIAIVFSGSAGALAGVVPQYGGGNGHGGNSGNSSELNIYQYGGNSALALQ 60  
QY 61 SDARKSETTITQSGYNGADVGQADNSTIELTQGFNNATIDQWNAKNSDITVQYGG 120  
DB 61 TDARNSDLTITQGGGNGADVGQSDSSIDLITQRFNGSATLDQWNGKNSMTVKQFGG 120  
QY 121 NNAALVNQTSADSSVMVQVGFNNATANQY 151  
DB 121 GNGAAVDQTASNSVNVTVQVGFNNATAHQY 151  
Search completed: August 2, 2004, 14:48:25  
Job time : 44.9 secs



Result No.	Score	Query		Length	DB	ID	Description
		Match	%				
1	686	87.9	151	1	US-08-233-788A-59	Sequence 59, Appl	
2	506	64.9	120	1	US-08-233-788A-57	Sequence 57, Appl	
3	92	11.8	975	4	US-09-328-352-4764	Sequence 4764, Ap	
4	90	11.5	906	1	US-08-254-573-2	Sequence 2, Appl	
5	90	11.5	906	1	US-08-687-379-2	Sequence 2, Appl	
6	90	11.5	906	1	US-08-687-379-4	Sequence 4, Appl	
7	90	11.5	906	4	US-08-172-332-1	Sequence 1, Appl	
8	90	11.5	906	4	US-08-216-326-2	Sequence 2, Appl	
9	87.5	11.2	738	3	US-08-864-038A-5	Sequence 3, Appl	
10	84	10.8	892	4	US-09-336-447A-3	Sequence 5, Appl	
11	84	10.8	907	1	US-07-718-575-2	Sequence 2, Appl	
12	84	10.8	907	1	US-08-481-206-2	Sequence 2, Appl	
13	84	10.8	907	2	US-08-486-269A-2	Sequence 2, Appl	
14	84	10.8	943	4	US-09-056-556-204	Sequence 204, App	
15	84	10.8	943	4	US-09-072-596-199	Sequence 199, App	
16	84	10.8	943	4	US-09-477-135A-131	Sequence 131, App	
17	84	10.8	943	4	US-09-072-967-204	Sequence 204, App	
18	84	10.8	1415	4	US-09-252-991A-26438	Sequence 26438, A	
19	83.5	10.7	415	4	US-09-025-769B-280	Sequence 280, App	
20	83.5	10.7	873	4	US-09-336-447A-13	Sequence 13, Appl	
21	83	10.6	518	3	US-09-043-123-2	Sequence 2, Appl	
22	83	10.6	1912	1	US-08-409-995-4	Sequence 4, Appl	
23	83	10.6	1912	3	US-08-685-467-4	Sequence 4, Appl	
24	83	10.6	2353	3	US-08-377-155-33	Sequence 33, Appl	
25	83	10.6	2353	3	US-08-913-942-4	Sequence 4, Appl	
26	83	10.6	2353	4	US-09-669-974-33	Sequence 33, Appl	
27	83	10.6	2353	4	US-09-797-862-33	Sequence 33, Appl	

Db 1 MLLKVAFAAIVVSGSALAGVVPQWGGGNGHNGSGSPDSTLSIYQVGSANAALALQ 60  
 QY 61 SDARKSETTITQSGYNGADVGGAANSTIELTQNGFRNNATIDQWNAKNSDITVGYGG 120  
 Db 61 SDARKSETTITQSGYNGADVGGAANSTIELTQNGFRNNATIDQWNAKNSDITVGYGG 120  
 QY 121 NNAALVNOTASDSSVVRQVGFNNATANY 151  
 Db 121 NNPALVNOTASDSSVVRQVGFNNATANY 151

## RESULT 2

US-08-233-788A-57  
 ; Sequence 57, Application US/08233788A  
 ; Patent No. 5635617

## ; GENERAL INFORMATION:

; APPLICANT: Doran, James L.  
 ; APPLICANT: Kay, William W.  
 ; APPLICANT: Collinson, Karen S.  
 ; APPLICANT: Clouthier, Sharon C.  
 ; TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR DETECTION  
 ; OF SALMONELLA  
 ; NUMBER OF SEQUENCES: 61  
 ; CORRESPONDENCE ADDRESS:  
 ; ADDRESSEE: Seed and Berry  
 ; STREET: 6300 Columbia Center, 701 Fifth Avenue  
 ; CITY: Seattle  
 ; STATE: Washington  
 ; COUNTRY: U.S.A.  
 ; ZIP: 98104-7092

## ; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk  
 ; COMPUTER: IBM PC compatible  
 ; OPERATING SYSTEM: PC-DOS/MS-DOS  
 ; SOFTWARE: Patent In Release #1.0, Version #1.25  
 ; CURRENT APPLICATION DATA: US/08/233,788A  
 ; FILING DATE: 26-APR-1994  
 ; CLASSIFICATION: 435  
 ; ATTORNEY/AGENT INFORMATION:  
 ; NAME: King, Joshua  
 ; REGISTRATION NUMBER: 35,570  
 ; REFERENCE/DOCKET NUMBER: 920043.403C2  
 ; TELECOMMUNICATION INFORMATION:  
 ; TELEPHONE: (206) 622-4900  
 ; TELEFAX: (206) 682-6031  
 ; TELEFAX: 3723836 SEEDANBERRY  
 ; INFORMATION FOR SEQ ID NO: 57:  
 ; SEQUENCE CHARACTERISTICS:  
 ; LENGTH: 120 amino acids  
 ; TYPE: amino acid  
 ; TOPOLOGY: linear  
 ; MOLECULE TYPE: protein

US-08-233-788A-57

Query Match 64.9%; Score 506; DB 1; Length 120;  
 Best Local Similarity 87.5%; Pred. No. 3.4e-44;  
 Matches 98; Conservative 3; Mismatches 11; Indels 0; Gaps 0;  
 QY 22 VVPQWGGGNGHNGSGSPDYDQLVTRVVTHEMAHALQSDARKSETTITQSGYNGADV 81  
 Db 1 VVPQWGGGNGHNGSGSPDSTLSIYQVGSANAALQSDARKSETTITQSGYNGADV 60  
 QY 82 GQGANSTIELTQNGFRNNATIDQWNAKNSDITVGYGNNALVNOTASDS 133  
 Db 61 GQGANSTIELTQNGFRNNATIDQWNAKNSDITVGYGNNALVNOTASDS 112

## RESULT 3

US-09-328-352-4764  
 ; Sequence 4764, Application US/09328352  
 ; Patent No. 6562958  
 ; GENERAL INFORMATION:

; APPLICANT: Gary L. Breton et al.  
 ; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO ACINETOBACTER  
 ; FILE REFERENCE: GTC99-03PA  
 ; CURRENT APPLICATION NUMBER: US/09/328,352  
 ; CURRENT FILING DATE: 1999-06-04  
 ; NUMBER OF SEQ ID NOS: 8252  
 ; SEQ ID NO 4764  
 ; LENGTH: 975  
 ; TYPE: PRT  
 ; ORGANISM: Acinetobacter baumannii  
 US-09-328-352-4764

Query Match 11.8%; Score 92; DB 4; Length 975;  
 Best Local Similarity 23.8%; Pred. No. 0.75;  
 Matches 35; Conservative 24; Mismatches 52; Indels 36; Gaps 7;

QY 15 SGSALAGVVPQWGGGNGHNGSGSPDYDQLVTRVVTHEMAHALQSDA 63  
 Db 300 AGNGIA-----SCNGEHNYGIGNGGDDVD--ITAPITGVLNISGNSFTLIGNSSSSSV 351  
 QY 64 RKSETTITQS-----GYGNGADVQGCADNSTIELTQNGF-----RNNATIDQWNAKNS 111  
 Db 352 NTAPTITNTVNDTIDNGSGGTGSGGSGDGLLNGAASNGEHNYGIGNGGDDV 411  
 QY 112 DIT-----VGYGNNALVNOTASDS 133  
 Db 412 DITSPITGIFNFCNSFSLIGNSSSS 438

## RESULT 4

US-08-254-573-2  
 ; Sequence 2, Application US/08254573  
 ; Patent No. 5610032

## ; GENERAL INFORMATION:

; APPLICANT: KAMBOJ, Rajender  
 ; APPLICANT: ELIOTT, Cardace  
 ; APPLICANT: NUTT, Stephen  
 ; TITLE OF INVENTION: AMPEA-BINDING HUMAN GIUR1 RECEPTORS  
 ; NUMBER OF SEQUENCES: 5  
 ; CORRESPONDENCE ADDRESS:  
 ; ADDRESSEE: Foley & Lardner  
 ; STREET: 1800 Diagonal Road, Suite 500  
 ; CITY: Alexandria  
 ; STATE: Virginia  
 ; COUNTRY: USA  
 ; ZIP: 22313-0299  
 ; COMPUTER READABLE FORM:  
 ; MEDIUM TYPE: Floppy disk  
 ; COMPUTER: IBM PC compatible  
 ; OPERATING SYSTEM: PC-DOS/MS-DOS  
 ; SOFTWARE: Patent In Release #1.0, Version #1.25  
 ; CURRENT APPLICATION DATA: US/08/254,573  
 ; FILING DATE: 06-JUN-1994  
 ; CLASSIFICATION: 435  
 ; PRIOR APPLICATION DATA:  
 ; APPLICATION NUMBER: US 07/896,611  
 ; FILING DATE: 10-JUN-1992  
 ; ATTORNEY/AGENT INFORMATION:  
 ; NAME: BENT, Stephen A.  
 ; REGISTRATION NUMBER: 29,768  
 ; REFERENCE/DOCKET NUMBER: 16777/179 ALLE  
 ; TELECOMMUNICATION INFORMATION:  
 ; TELEPHONE: (703) 836-9300  
 ; TELEFAX: (703) 683-4109  
 ; TELEX: 899149

INFORMATION FOR SEQ ID NO: 2:  
 SEQUENCE CHARACTERISTICS:  
 LENGTH: 906 amino acids  
 TYPE: amino acid  
 TOPOLOGY: linear  
 MOLECULE TYPE: protein

```

US-08-254-573-2
Query Match      11.5%; Score 90; DB 1; Length 906;
Best Local Similarity 22.6%; Pred. No. 1.1;
Matches 45; Conservative 26; Mismatches 56; Indels 72; Gaps 10;

QY 6 VAAPAAIVVSGSALAGVVPQWGGGNGHNGSGGPDYDQ-----VTRVVTHE-----M 55
DB 250 VTGFQLVNYTDTIPAKIMQW-----KNSDARDHTRVDWKRPKYTSALTVDGVKVM 300

QY 56 AHALQSDARKSETTITQSGYNGADV-----GQGAD-----NSTIELTQ 94
DB 301 AEAFQSLRRQ---RIDISRRGNAGDCLANPAVWPWGQIDIQALQOVRFEGLTGNVQFNE 357

QY 95 NGRFNAT-----IDWNNAKNSDI---TVGYGNNALVNQT-----ASD 132
DB 358 KGRRTNYTLHVTEMKHDGIRKIGYWNEDDKFVPAATDAQAGDSSVQNRITYIVTTILED 417

QY 133 SSVVVRQVGFNNATANQY 151
DB 418 PYVMLKK-----NANQF 429

RESULT 5
US-08-687-379-2
; Sequence 2, Application US/08687379
; Patent No. 5756697
; GENERAL INFORMATION:
; APPLICANT: Hoeger, Thomas
; APPLICANT: Ulsch, Andreas
; APPLICANT: Bach, Alfred
; APPLICANT: Sterrer, Sylvia
; APPLICANT: Lemaitre, Hans-Georg
; TITLE OF INVENTION: Subunits of Glutamate Receptors, Their
; TITLE OF INVENTION: Preparation and Their Use
; NUMBER OF SEQUENCES: 26
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Keil & Weinlauf
; STREET: 1101 Connecticut Avenue
; CITY: Washington
; STATE: D.C.
; COUNTRY: USA
; ZIP: 20036
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette-3.5 inch, 720 Kb storage
; COMPUTER: IBM AT-compatible, 80286 processor
; OPERATING SYSTEM: MS-DOS version 6.0
; SOFTWARE: Wordperfect version 5.1
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/687,379
; FILING DATE: 05-AUG-1996
; CLASSIFICATION: 435
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 906 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-687-379-2

Query Match      11.5%; Score 90; DB 1; Length 906;
Best Local Similarity 22.6%; Pred. No. 1.1;
Matches 45; Conservative 26; Mismatches 56; Indels 72; Gaps 10;

QY 6 VAAPAAIVVSGSALAGVVPQWGGGNGHNGSGGPDYDQ-----VTRVVTHE-----M 55
DB 250 VTGFQLVNYTDTIPAKIMQW-----KNSDARDHTRVDWKRPKYTSALTVDGVKVM 300

QY 56 AHALQSDARKSETTITQSGYNGADV-----GQGAD-----NSTIELTQ 94
DB 301 AEAFQSLRRQ---RIDISRRGNAGDCLANPAVWPWGQIDIQALQOVRFEGLTGNVQFNE 357

QY 95 NGRFNAT-----IDWNNAKNSDI---TVGYGNNALVNQT-----ASD 132
DB 358 KGRRTNYTLHVTEMKHDGIRKIGYWNEDDKFVPAATDAQAGDSSVQNRITYIVTTILED 417

QY 133 SSVVVRQVGFNNATANQY 151
DB 418 PYVMLKK-----NANQF 429

RESULT 6
US-08-687-379-4
; Sequence 4, Application US/08687379
; Patent No. 5756697
; GENERAL INFORMATION:
; APPLICANT: Hoeger, Thomas
; APPLICANT: Ulsch, Andreas
; APPLICANT: Bach, Alfred
; APPLICANT: Sterrer, Sylvia
; APPLICANT: Lemaitre, Hans-Georg
; TITLE OF INVENTION: Subunits of Glutamate Receptors, Their
; TITLE OF INVENTION: Preparation and Their Use
; NUMBER OF SEQUENCES: 26
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Keil & Weinlauf
; STREET: 1101 Connecticut Avenue
; CITY: Washington
; STATE: D.C.
; COUNTRY: USA
; ZIP: 20036
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette-3.5 inch, 720 Kb storage
; COMPUTER: IBM AT-compatible, 80286 processor
; OPERATING SYSTEM: MS-DOS version 6.0
; SOFTWARE: Wordperfect version 5.1
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/687,379
; FILING DATE: 05-AUG-1996
; CLASSIFICATION: 435
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 906 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-687-379-4

Query Match      11.5%; Score 90; DB 1; Length 906;
Best Local Similarity 22.6%; Pred. No. 1.1;
Matches 45; Conservative 26; Mismatches 56; Indels 72; Gaps 10;

QY 6 VAAPAAIVVSGSALAGVVPQWGGGNGHNGSGGPDYDQ-----VTRVVTHE-----M 55
DB 250 VTGFQLVNYTDTIPAKIMQW-----KNSDARDHTRVDWKRPKYTSALTVDGVKVM 300

QY 56 AHALQSDARKSETTITQSGYNGADV-----GQGAD-----NSTIELTQ 94
DB 301 AEAFQSLRRQ---RIDISRRGNAGDCLANPAVWPWGQIDIQALQOVRFEGLTGNVQFNE 357

QY 95 NGRFNAT-----IDWNNAKNSDI---TVGYGNNALVNQT-----ASD 132
DB 358 KGRRTNYTLHVTEMKHDGIRKIGYWNEDDKFVPAATDAQAGDSSVQNRITYIVTTILED 417

QY 133 SSVVVRQVGFNNATANQY 151
DB 418 PYVMLKK-----NANQF 429

RESULT 7
US-08-172-332-1
; Sequence 1, Application US/08172332
; Patent No. 6313279
; GENERAL INFORMATION:
; APPLICANT: Burnett, J. Paul
; APPLICANT: Mayne, Nancy G

```

APPLICANT: Snyder, Yvonne M  
TITLE OF INVENTION: HUMAN GLUTAMATE RECEPTOR AND RELATED DNA  
TITLE OF INVENTION: COMPOUNDS  
NUMBER OF SEQUENCES: 5  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Lilly Patent Division/JPL  
STREET: Lilly Corporate Center  
CITY: Indianapolis  
STATE: IN  
COUNTRY: USA  
ZIP: 46285  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/172,332  
FILING DATE: 22-DEC-1993  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 07/879,688  
FILING DATE: May 1, 1992  
ATTORNEY/AGENT INFORMATION:  
NAME: Leeds, James P.  
REGISTRATION NUMBER: 35241  
REFERENCE/DOCKET NUMBER: X-8342  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 317-276-1667  
INFORMATION FOR SEQ ID NO: 1:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 906 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-172-332-1

Query Match 11.5%; Score 90; DB 4; Length 906;  
Best Local Similarity 22.6%; Pred. No. 1.1;  
Matches 45; Conservative 26; Mismatches 56; Indels 72; Gaps 10;  
QY 6 VAAFAIIVSGSALAGVVPQWGGGNGHNGSSGGPDYDQ-----VTRVVTHE-----M 55  
DB 250 VTGFQVLYNYDTTPAKIMQW-----KNSDARDHTRVDWKRPKYTSALTYDGVKVM 300  
QY 56 AHALQSDARKSETTITQSGYNGADV-----GQAD-----NSTIELTQ 94  
DB 301 AEAFQSLRRQ---RIDISRRGNAGDCLANPAVPWGGQIDIQALQQVRFEGLTGNVQFNE 357  
QY 95 NGRFNAT-----IDWNAKNSDI---TVGQYGGNNAALVNOT-----ASD 132  
DB 358 KGRNTYTLHVIEMKHGDKIRKIGYWNEDDKFVPAATDAQAGGDNSSVQNETYIVTTILED 417  
QY 133 SSVWVQVGFNGNATANQY 151  
DB 418 PYVWLKK-----NANQF 429

RESULT 8  
US-08-216-326-2  
Sequence 2, Application US/08216326  
Patent No. 6406868  
GENERAL INFORMATION:  
APPLICANT: KAMEOJ, Rajender  
APPLICANT: ELIOTT, Candace  
TITLE OF INVENTION: AMPA-BINDING HUMAN GLUR1 RECEPTORS  
NUMBER OF SEQUENCES: 4  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Foley & Lardner  
STREET: 3000 K Street N.W., Suite 500  
CITY: Washington, D.C.

COUNTRY: USA  
ZIP: 20007-5109  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/216,326  
FILING DATE: 23-MAR-1994  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/896,611  
FILING DATE: 10-JUN-1992  
ATTORNEY/AGENT INFORMATION:  
NAME: BENT, Stephen A.  
REGISTRATION NUMBER: 29,768  
REFERENCE/DOCKET NUMBER: 16777/229/ALLE  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (202) 672-5300  
TELEFAX: (202) 672-5399  
TELEX: 904136  
INFORMATION FOR SEQ ID NO: 2:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 906 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-216-326-2

Query Match 11.5%; Score 90; DB 4; Length 906;  
Best Local Similarity 22.6%; Pred. No. 1.1;  
Matches 45; Conservative 26; Mismatches 56; Indels 72; Gaps 10;  
QY 6 VAAFAIIVSGSALAGVVPQWGGGNGHNGSSGGPDYDQ-----VTRVVTHE-----M 55  
DB 250 VTGFQVLYNYDTTPAKIMQW-----KNSDARDHTRVDWKRPKYTSALTYDGVKVM 300  
QY 56 AHALQSDARKSETTITQSGYNGADV-----GQAD-----NSTIELTQ 94  
DB 301 AEAFQSLRRQ---RIDISRRGNAGDCLANPAVPWGGQIDIQALQQVRFEGLTGNVQFNE 357  
QY 95 NGRFNAT-----IDWNAKNSDI---TVGQYGGNNAALVNOT-----ASD 132  
DB 358 KGRNTYTLHVIEMKHGDKIRKIGYWNEDDKFVPAATDAQAGGDNSSVQNETYIVTTILED 417  
QY 133 SSVWVQVGFNGNATANQY 151  
DB 418 PYVWLKK-----NANQF 429

RESULT 9  
US-08-864-038A-3  
Sequence 3, Application US/08864038A  
Patent No. 6001592  
GENERAL INFORMATION:  
APPLICANT: Kunio NAKASHIMA et al.  
TITLE OF INVENTION: NOVEL POLYPEPTIDE GENE CDNA, VECTOR  
CONTAINING SAID CDNA, HOST CELLS TRANSFORMED WITH SAID  
TITLE OF INVENTION: VECTOR, POLYPEPTIDE PRODUCED THEREBY, METHOD OF PRODUCING  
TITLE OF INVENTION: SAID POLYPEPTIDE, DNA ENCODING SAID POLYPEPTIDE AND ANTIBODY  
TITLE OF INVENTION: TO SAID POLYPEPTIDE  
NUMBER OF SEQUENCES: 4  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: 812-5 Hirano  
STREET: Isshinden  
CITY: Tsu-city  
STATE: Mie-prefecture  
COUNTRY: JAPAN  
ZIP: 514-01  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette, 3.50 inch, 1.44 MB storage  
COMPUTER: IBM Compatible  
OPERATING SYSTEM: Microsoft Windows 95

```

; SOFTWARE: Word Perfect 6.1
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/864,038A
; FILING DATE: May 28, 1997
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: JP 8-184459
; FILING DATE: 15-July-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: C. Bruce Hamburg
; REGISTRATION NUMBER: 22,389
; REFERENCE/DOCKET NUMBER: F-5610
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212)986-2340
; TELEFAX: (212)953-7733
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 738
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; ORIGINAL SOURCE:
; ORGANISM: Pinctada fucata
; CELL TYPE: mantle epithelial cell
; FEATURE:
; NAME/KEY: peptide
; LOCATION: from 1 to 738
; IDENTIFICATION METHOD: E (by experiment)
; US-08-864-038A-3

Query Match 11.2%; Score 87.5; DB 3; Length 738;
Best Local Similarity 25.6%; Pred. No. 1.5; Indels 49; Gaps 5;
Matches 41; Conservative 11; Mismatches 59;

QY 3 LKVAFAAIVVSGSALAGVVPQWGGNGHNGGSSGPDYDQLVTRVVTHMAHALOSD 62
DB 419 LKSSASASASASASAG-----GGGGGGGGGGGGGG-----GG 455
QY 63 ARKSETTITQSGYNGADVQGGADNSTIELTQ-----NGFRNATIDQNAKNSDITV 117
DB 456 AGAALAAALAAAGAGGLGGGGGGGALAAALAAAGAGGGGFGGLGGL-----GG 503
QY 118 YGNNALVNOTASDS-----VMVRQVGFNNATA 148
DB 504 LGGSAALAAAAAASAGGGGGRALRALRQVREGGSA 543

RESULT 10
US-09-336-447A-5
; Sequence 5, Application US/09336447A
; Patent No. 6310190
; GENERAL INFORMATION:
; APPLICANT: HANSEN, ERIC J.
; APPLICANT: ABEI, CHRISTOPH
; APPLICANT: COPE, LESLIE D.
; APPLICANT: MACIVER, ISOBEL
; APPLICANT: FISKE, MICHAEL J.
; APPLICANT: FREDENBURG, ROSS A.
; TITLE OF INVENTION: USP1 AND USP2 ANTIGENS OF MORAXELLA CATARRHALIS
; FILE REFERENCE: ANCY.024
; CURRENT APPLICATION NUMBER: US/09/336,447A
; CURRENT FILING DATE: 1999-06-21
; NUMBER OF SEQ ID NOS: 98
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 5
; LENGTH: 892
; TYPE: PRT
; ORGANISM: Moraxella catarrhalis
; US-09-336-447A-5

Query Match 10.8%; Score 84; DB 4; Length 892;
Best Local Similarity 26.8%; Pred. No. 4.3; Indels 55; Gaps 9;
Matches 41; Conservative 17; Mismatches 55;

QY 118 YGNNALVNOTASDS-----VMVRQVGFNNATA 148
DB 504 LGGSAALAAAAAASAGGGGGRALRALRQVREGGSA 543
```

```

QY 28 GGGNHNGGNS--GPDYDQLVTRVVTHMAHALOSDARKSETTI-----TQSGYNGAD 80
DB 75 GKDNKAGNYSVVGDDYNEAKGNYST--VGGSSNTAKGKSTIGGGSDTNDANGTYST 132
QY 81 VGGQ-----ADNSTI-----ELTQNGFRNATIDQNAKNSDITVGYG--G 120
DB 133 IGGYYSRAIGDSSTIGGGYNOATGKSTVAGRRN-----QATGNNSTVAGSYNQATG 188
QY 121 NNAALV-----NOTASDSSVMVRQVGFNNATAN 149
DB 189 NNSTVAGGSHNOATGEGSF--AAGVENKANAN 218

RESULT 11
US-07-718-575-2
; Sequence 2, Application US/07718575
; Patent No. 5202257
; GENERAL INFORMATION:
; APPLICANT: Heinemann Ph.D., Stephen F.
; APPLICANT: Boulter Ph.D., James R.
; APPLICANT: Hollmann Ph.D., Michael NMN
; APPLICANT: Bettler Ph.D., Bernhard NMN
; APPLICANT: Jensen Ph.D., Jan E.
; TITLE OF INVENTION: GLUTAMATE RECEPTOR COMPOSITIONS AND
; METHOD OF INVENTION: METHODS
; NUMBER OF SEQUENCES: 14
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Pretty, Schroeder, Brueggemann & Clark
; STREET: 444 So. Flower St., Suite 2000
; CITY: Los Angeles
; STATE: California
; COUNTRY: United States
; ZIP: 90071-2921
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/718,575
; FILING DATE: 19910813
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Reiter Ph.D., Stephen E.
; REGISTRATION NUMBER: 31192
; REFERENCE/DOCKET NUMBER: P31 8962
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (619) 535-9001
; TELEFAX: (619) 535-8949
; TELEX: 9103330318
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 907 amino acids
; TYPE: AMINO ACID
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-07-718-575-2

Query Match 10.8%; Score 84; DB 1; Length 907;
Best Local Similarity 22.6%; Pred. No. 4.4; Indels 54; Gaps 8;
Matches 43; Conservative 24; Mismatches 69;

QY 6 VAAFAAIVVSGSALAGVVPQWGGNGHNGG--NSGGPDYDQLVTRVVTHMAHALOSDAR 64
DB 250 VTGFOLVNYTDTIPARIMQWRTSDSRDTRVDWKRPKYTSALTVDGVKVMFAFQSLRR 309
QY 65 KSETTITQSGYNGADV-----GQAD-----NSTIELTQNGFRNNAT- 102
DB 310 Q---RIDISRRGNAGDCLANFAVPGQIDTQALQVRFEGTLGNVQFNEKGRRTNYTL 366
QY 103 -----IDQNAKNSDI--TVQYQVGGNNAALVNOT-----ASDSSVMVRQV 141
DB 367 HVIEKMDGIRKIGYWNEDDKFVPAATDAQAGGDNSSVQNRITYIVTTILEDPYVMLK-- 424
```

QY 142 FGNNATANQY 151  
Db 425 -----NANQF 429

RESULT 12  
US-08-481-206-2  
; Sequence 2, Application US/08481206  
; Patent No. 5739291  
; GENERAL INFORMATION:  
; APPLICANT: Heinemann Ph.D., Stephen F.  
; APPLICANT: Boulter Ph.D., James R.  
; APPLICANT: Hollmann Ph.D., Michael NMN  
; APPLICANT: Bettler Ph.D., Bernhard NMN  
; APPLICANT: Jensen Ph.D., Jan E.  
; TITLE OF INVENTION: GLUTAMATE RECEPTOR COMPOSITIONS AND  
; TITLE OF INVENTION: METHODS  
; NUMBER OF SEQUENCES: 14  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Pretty, Schroeder, Brueggemann & Clark  
; STREET: 444 So. Flower St., Suite 2000  
; CITY: Los Angeles  
; STATE: California  
; COUNTRY: United States  
; ZIP: 90071-2921  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patent In Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/481,206  
; FILING DATE:  
; CLASSIFICATION: 530  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US/08/013,767  
; FILING DATE: 04-FEB-1993  
; APPLICATION NUMBER: 07/718,575  
; FILING DATE: 21-JUN-1991  
; APPLICATION NUMBER: PCT/US90/06153  
; FILING DATE: 25-OCT-1990  
; APPLICATION NUMBER: 07/428,116  
; FILING DATE: 27-OCT-1989  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Reiter, Stephen E.  
; REGISTRATION NUMBER: 31192  
; REFERENCE/DOCKET NUMBER: P31 8962  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (619) 535-9001  
; TELEFAX: (619) 535-8949  
; TELEX: 9103330318  
; INFORMATION FOR SEQ ID NO: 2:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 907 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
US-08-481-206-2

Query Match 10.8%; Score 84; DB 1; Length 907;  
Best Local Similarity 22.6%; Pred. No. 4.4;  
Matches 43; Conservative 24; Mismatches 69; Indels 54; Gaps 8;

QY 6 VAAFAAIVVSGSALAGVWPQGGGNGHGGG-NSSGPDYDQLVTRVVTHEMAHALQSDAR 64  
Db 250 VTGFLVNYTDTIPARIMQWRTSDSRDTRVDWKPKYTSALTVDGKVMAEAFQSLRR 309  
QY 65 KSETTTTQSGYNGADV-----CQGD-----NSTIETQNGFRNAT- 102  
Db 310 Q-----RIDISRRGNAGCLANPAVFWGQGDIDQALQVRFEGLTGNVQNEKGRNTYL 366  
QY 103 -----IDONNAKNSDI---TVGOYGNNALVNOT-----ASDSSVMVRQVG 141  
Db 367 HVIEKHDGIRKIGYWNEDDKFVPAATDAQAGDSSVQNRYYIVTTILEDPPYVMLK-- 424  
QY 142 FGNNATANQY 151  
Db 425 -----NANQF 429

RESULT 13  
US-08-486-269A-2  
; Sequence 2, Application US/08486269A  
; Patent No. 5945509  
; GENERAL INFORMATION:  
; APPLICANT: Heinemann, Stephen F.  
; APPLICANT: Boulter, James R.  
; APPLICANT: Hollmann, Michael  
; APPLICANT: Bettler, Bernhard  
; APPLICANT: Jensen, Jan E.  
; TITLE OF INVENTION: GLUTAMATE RECEPTOR COMPOSITIONS  
; TITLE OF INVENTION: AND METHODS  
; NUMBER OF SEQUENCES: 14  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Gray Cary Ware & Freidenrich LLP  
; STREET: 4365 Executive Drive, Suite 1600  
; CITY: San Diego  
; STATE: CA  
; COUNTRY: USA  
; ZIP: 92121  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Diskette  
; COMPUTER: IBM Compatible  
; OPERATING SYSTEM: DOS  
; SOFTWARE: FastSeq for Windows DEMONSTRATION Version 2.0D  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/486,269A  
; FILING DATE: 06-JUN-1995  
; CLASSIFICATION: 530  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: 08/013,767  
; FILING DATE: 04-FEB-1993  
; APPLICATION NUMBER: 07/718,575  
; FILING DATE: 21-JUN-1991  
; APPLICATION NUMBER: PCT/US90/06153  
; FILING DATE: 25-OCT-1990  
; APPLICATION NUMBER: 07/428,116  
; FILING DATE: 27-OCT-1989  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Reiter, Stephen E.  
; REGISTRATION NUMBER: 31,192  
; REFERENCE/DOCKET NUMBER: P41 9986  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 619-677-1409  
; TELEFAX: 619-677-1465  
; TELEX:  
; INFORMATION FOR SEQ ID NO: 2:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 907 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
US-08-486-269A-2

Query Match 10.8%; Score 84; DB 2; Length 907;  
Best Local Similarity 22.6%; Pred. No. 4.4;  
Matches 43; Conservative 24; Mismatches 69; Indels 54; Gaps 8;

QY 6 VAAFAAIVVSGSALAGVWPQGGGNGHGGG-NSSGPDYDQLVTRVVTHEMAHALQSDAR 64  
Db 250 VTGFLVNYTDTIPARIMQWRTSDSRDTRVDWKPKYTSALTVDGKVMAEAFQSLRR 309  
QY 65 KSETTTTQSGYNGADV-----CQGD-----NSTIETQNGFRNAT- 102  
Db 310 Q-----RIDISRRGNAGCLANPAVFWGQGDIDQALQVRFEGLTGNVQNEKGRNTYL 366  
QY 103 -----IDONNAKNSDI---TVGOYGNNALVNOT-----ASDSSVMVRQVG 141  
Db 367 HVIEKHDGIRKIGYWNEDDKFVPAATDAQAGDSSVQNRYYIVTTILEDPPYVMLK-- 424



APPLICANT: Vedvick, Thomas S.  
APPLICANT: Twardzik, Daniel R.  
APPLICANT: Lodes, Michael J.  
APPLICANT: Hendrickson, Ronald C.  
TITLE OF INVENTION: COMPOUNDS AND METHODS FOR DIAGNOSIS OF TUBERCULOSIS  
NUMBER OF SEQUENCES: 350  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: SEED and BERRY LLP  
STREET: 6300 Columbia Center, 701 Fifth Avenue  
CITY: Seattle  
STATE: Washington  
COUNTRY: USA  
ZIP: 98104-7092  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/072,596  
FILING DATE: 05-MAY-1998  
CLASSIFICATION:  
ATTORNEY/AGENT INFORMATION:  
NAME: Maki, David J.  
REGISTRATION NUMBER: 31,392  
REFERENCE/DOCKET NUMBER: 210121.417C9  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (206) 622-4900  
TELEFAX: (206) 682-6031  
INFORMATION FOR SEQ ID NO: 199:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 943 amino acids  
TYPE: amino acid  
STRANDEDNESS:  
TOPOLOGY: linear  
US-09-072-596-199

Query Match 10.8%; Score 84; DB 4; Length 943;  
Best Local Similarity 25.2%; Pred. No. 4.7;  
Matches 31; Conservative 9; Mismatches 41; Indels 42; Gaps 4;

QY 15 SGSALAGVVPQWGGGHHNGGSSGPDYDOLVTRVVTHEMAHALQSDARKSETTITQSG 74  
Db 184 SGTGNVGI-----GNSGTGNWGNIGNSNSYN-----TG 211

QY 75 YNGADVGQGDNSTIELTQNGFRNATTIDWNNAKNSD---ITVGQY-----GGNNA 124  
Db 212 FGNSGDANTGFFNSGIANTGVNAGNYTGSYNGNSNTGCFNMGOYNTGNSNYNTG 271

QY 125 LVN 127  
Db 272 LAN 274

Search completed: August 2, 2004, 14:58:33  
Job time : 13 secs

APPLICANT: Vedvick, Thomas S.  
APPLICANT: Twardzik, Daniel R.  
APPLICANT: Lodes, Michael J.  
APPLICANT: Hendrickson, Ronald C.  
TITLE OF INVENTION: COMPOUNDS AND METHODS FOR DIAGNOSIS OF TUBERCULOSIS  
NUMBER OF SEQUENCES: 350  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: SEED and BERRY LLP  
STREET: 6300 Columbia Center, 701 Fifth Avenue  
CITY: Seattle  
STATE: Washington  
COUNTRY: USA  
ZIP: 98104-7092  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/056,556  
FILING DATE: 07-APR-1998  
CLASSIFICATION:  
ATTORNEY/AGENT INFORMATION:  
NAME: Maki, David J.  
REGISTRATION NUMBER: 31,392  
REFERENCE/DOCKET NUMBER: 210121.457  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (206) 622-4900  
TELEFAX: (206) 682-6031  
INFORMATION FOR SEQ ID NO: 204:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 943 amino acids  
TYPE: amino acid  
STRANDEDNESS:  
TOPOLOGY: linear  
US-09-056-556-204

Query Match 10.8%; Score 84; DB 4; Length 943;  
Best Local Similarity 25.2%; Pred. No. 4.7;  
Matches 31; Conservative 9; Mismatches 41; Indels 42; Gaps 4;

QY 15 SGSALAGVVPQWGGGHHNGGSSGPDYDOLVTRVVTHEMAHALQSDARKSETTITQSG 74  
Db 184 SGTGNVGI-----GNSGTGNWGNIGNSNSYN-----TG 211

QY 75 YNGADVGQGDNSTIELTQNGFRNATTIDWNNAKNSD---ITVGQY-----GGNNA 124  
Db 212 FGNSGDANTGFFNSGIANTGVNAGNYTGSYNGNSNTGCFNMGOYNTGNSNYNTG 271

QY 125 LVN 127  
Db 272 LAN 274

Search completed: August 2, 2004, 14:58:33  
Job time : 13 secs

RESULT 14  
US-09-056-556-204  
Sequence 204, Application US/09056556  
Patent No. 6350456  
GENERAL INFORMATION:  
APPLICANT: Reed, Steven G.  
APPLICANT: Skeiky, Yasir A.W.  
APPLICANT: Dillon, Davin C.  
TITLE OF INVENTION: COMPOUNDS AND METHODS FOR THE PREVENTION AND TREAT  
NUMBER OF SEQUENCES: 241  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: SEED and BERRY LLP  
STREET: 6300 Columbia Center, 701 Fifth Avenue  
CITY: Seattle  
STATE: Washington  
COUNTRY: USA  
ZIP: 98104-7092  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/056,556  
FILING DATE: 07-APR-1998  
CLASSIFICATION:  
ATTORNEY/AGENT INFORMATION:  
NAME: Maki, David J.  
REGISTRATION NUMBER: 31,392  
REFERENCE/DOCKET NUMBER: 210121.457  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (206) 622-4900  
TELEFAX: (206) 682-6031  
INFORMATION FOR SEQ ID NO: 204:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 943 amino acids  
TYPE: amino acid  
STRANDEDNESS:  
TOPOLOGY: linear  
US-09-056-556-204

Query Match 10.8%; Score 84; DB 4; Length 943;  
Best Local Similarity 25.2%; Pred. No. 4.7;  
Matches 31; Conservative 9; Mismatches 41; Indels 42; Gaps 4;

QY 15 SGSALAGVVPQWGGGHHNGGSSGPDYDOLVTRVVTHEMAHALQSDARKSETTITQSG 74  
Db 184 SGTGNVGI-----GNSGTGNWGNIGNSNSYN-----TG 211

QY 75 YNGADVGQGDNSTIELTQNGFRNATTIDWNNAKNSD---ITVGQY-----GGNNA 124  
Db 212 FGNSGDANTGFFNSGIANTGVNAGNYTGSYNGNSNTGCFNMGOYNTGNSNYNTG 271

QY 125 LVN 127  
Db 272 LAN 274

Search completed: August 2, 2004, 14:58:33  
Job time : 13 secs

RESULT 15  
US-09-072-596-199  
Sequence 199, Application US/09072596  
Patent No. 6458366  
GENERAL INFORMATION:  
APPLICANT: Reed, Steven G.  
APPLICANT: Skeiky, Yasir A.W.  
APPLICANT: Dillon, Davin C.  
APPLICANT: Campos-Neto, Antonia  
APPLICANT: Houghton, Raymond



GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: August 2, 2004, 14:54:48 ; Search time 36.8 Seconds  
(without alignments)  
1287.123 Million cell updates/sec

Title: US-09-543-407-18  
Perfect score: 780  
Sequence: 1 MKLLKVAFAAIVSGSALA.....DSSVMRVQVGFNNATANQY 151

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 1291235 seqs, 313682936 residues

Total number of hits satisfying chosen parameters: 1291235

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Published Applications AA:

- 1: /cgn2\_6/prodata/2/pubpaa/US07\_PUBCOMB.pcp.\*
- 2: /cgn2\_6/prodata/2/pubpaa/PCT\_NEW\_PUB.pcp.\*
- 3: /cgn2\_6/prodata/2/pubpaa/US06\_NEW\_PUB.pcp.\*
- 4: /cgn2\_6/prodata/2/pubpaa/US06\_PUBCOMB.pcp.\*
- 5: /cgn2\_6/prodata/2/pubpaa/US07\_NEW\_PUB.pcp.\*
- 6: /cgn2\_6/prodata/2/pubpaa/PCTUS\_PUBCOMB.pcp.\*
- 7: /cgn2\_6/prodata/2/pubpaa/US08\_NEW\_PUB.pcp.\*
- 8: /cgn2\_6/prodata/2/pubpaa/US09\_PUBCOMB.pcp.\*
- 9: /cgn2\_6/prodata/2/pubpaa/US09A\_PUBCOMB.pcp.\*
- 10: /cgn2\_6/prodata/2/pubpaa/US09B\_PUBCOMB.pcp.\*
- 11: /cgn2\_6/prodata/2/pubpaa/US09C\_PUBCOMB.pcp.\*
- 12: /cgn2\_6/prodata/2/pubpaa/US09\_NEW\_PUB.pcp.\*
- 13: /cgn2\_6/prodata/2/pubpaa/US10A\_PUBCOMB.pcp.\*
- 14: /cgn2\_6/prodata/2/pubpaa/US10B\_PUBCOMB.pcp.\*
- 15: /cgn2\_6/prodata/2/pubpaa/US10C\_PUBCOMB.pcp.\*
- 16: /cgn2\_6/prodata/2/pubpaa/US10\_NEW\_PUB.pcp.\*
- 17: /cgn2\_6/prodata/2/pubpaa/US60\_NEW\_PUB.pcp.\*
- 18: /cgn2\_6/prodata/2/pubpaa/US60\_PUBCOMB.pcp.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	520	66.7	151	US-09-741-873B-4	Sequence 4, Appli
2	520	66.7	151	US-09-741-873B-4	Sequence 4, Appli
3	444	56.9	131	US-09-741-873B-2	Sequence 2, Appli
4	444	56.9	131	US-09-741-873B-2	Sequence 2, Appli
5	108.5	13.9	445	US-10-369-493-20638	Sequence 20638, A
6	93	11.9	974	US-10-282-122A-44999	Sequence 4, Appli
7	91.5	11.7	580	US-10-647-057-4	Sequence 146, App
8	91	11.7	597	US-09-793-306-146	Sequence 7, Appli
9	90	11.5	906	US-10-233-449-7	Sequence 2, Appli
10	90	11.5	906	US-10-251-661-2	Sequence 998, App
11	89	11.4	1448	US-10-408-765A-998	Sequence 203972,
12	88.5	11.3	204	US-10-424-599-203972	Sequence 37715, A
13	88.5	11.3	244	US-10-425-114-37715	Sequence 114193,
14	88.5	11.3	253	US-10-437-963-114193	Sequence 232271,
15	88.5	11.3	283	US-10-424-599-232271	

16	88	11.3	3619	12	US-10-282-122A-67768	Sequence 67768, A
17	88	11.3	6310	12	US-10-282-122A-67793	Sequence 67793, A
18	87.5	11.2	172	12	US-10-425-114-45013	Sequence 45013, A
19	86.5	11.1	208	16	US-10-475-970-16	Sequence 16, Appl
20	85.5	11.0	278	9	US-09-810-264-28	Sequence 28, Appl
21	85.5	11.0	588	12	US-10-282-122A-64869	Sequence 64869, A
22	84.5	10.8	408	15	US-10-369-493-12833	Sequence 12833, A
23	84.5	10.8	847	16	US-10-437-963-118741	Sequence 118741, A
24	84	10.8	283	12	US-10-424-599-168890	Sequence 168890, A
25	84	10.8	892	10	US-09-952-267-5	Sequence 5, Appli
26	84	10.8	943	9	US-09-996-634-131	Sequence 131, App
27	84	10.8	943	10	US-09-997-182-131	Sequence 131, App
28	84	10.8	943	10	US-09-997-181-131	Sequence 131, App
29	84	10.8	943	14	US-10-193-002-199	Sequence 199, App
30	84	10.8	943	14	US-10-084-843-204	Sequence 204, App
31	84	10.8	1246	12	US-10-282-122A-49773	Sequence 49773, A
32	84	10.8	3300	12	US-10-282-122A-64369	Sequence 64369, A
33	83.5	10.7	65	9	US-09-996-194-16	Sequence 16, Appl
34	83.5	10.7	65	12	US-10-164-966-33	Sequence 33, Appl
35	83.5	10.7	628	12	US-10-282-122A-53269	Sequence 53269, A
36	83.5	10.7	873	10	US-09-952-267-13	Sequence 13, Appl
37	83	10.6	182	14	US-10-238-075-1549	Sequence 1549, Ap
38	83	10.6	354	10	US-09-820-843A-21	Sequence 21, Appl
39	83	10.6	518	9	US-09-976-297-2	Sequence 2, Appli
40	83	10.6	720	9	US-09-801-368-176	Sequence 176, App
41	83	10.6	961	12	US-10-282-122A-76322	Sequence 76322, A
42	83	10.6	1306	12	US-10-282-122A-64405	Sequence 64405, A
43	83	10.6	2353	9	US-09-797-862-33	Sequence 33, Appl
44	82.5	10.6	434	15	US-10-045-674-594	Sequence 594, App
45	82.5	10.6	574	12	US-10-282-122A-55429	Sequence 55429, A

ALIGNMENTS

RESULT 1

US-09-741-873B-4  
; Sequence 4, Application US/09741873B  
; Publication No. US20020081722A1  
; GENERAL INFORMATION:  
; APPLICANT: Normark, Staffan  
; APPLICANT: Olsen, Arne  
; TITLE OF INVENTION: Fibronectin Binding Protein As Well As Its Preparation  
; FILE REFERENCE: 012889-084  
; CURRENT FILING DATE: 2003-04-04  
; PRIOR APPLICATION NUMBER: SE 8801723-1  
; PRIOR FILING DATE: 1998-05-06  
; PRIOR APPLICATION NUMBER: US 08/978,878  
; PRIOR FILING DATE: 1997-11-26  
; PRIOR APPLICATION NUMBER: US 07/347,189  
; PRIOR FILING DATE: 1989-05-04  
; PRIOR APPLICATION NUMBER: US 07/789,437  
; PRIOR FILING DATE: 1991-11-06  
; PRIOR APPLICATION NUMBER: US 07/970,846  
; PRIOR FILING DATE: 1992-11-03  
; PRIOR APPLICATION NUMBER: US 08/187,865  
; PRIOR FILING DATE: 1994-01-28  
; PRIOR APPLICATION NUMBER: US 08/318,519  
; PRIOR FILING DATE: 1994-10-05  
; NUMBER OF SEQ ID NOS: 10  
; SOFTWARE: PatentIn version 3.0  
; SEQ ID NO 4  
; LENGTH: 151  
; TYPE: PRT  
; ORGANISM: Escherichia coli  
US-09-741-873B-4

Query Match 66.7%; Score 520; DB 12; Length 151;  
Best Local Similarity 68.2%; Pred. No. 2.2e-45;  
Matches 103; Conservative 21; Mismatches 27; Indels 0; Gaps 0;

QY 1 MKLLKVAFAAIVSGSALLAGVVPQWGGGNGSGGPDYDQLVIRVVTHEMAHALQ 60

```

Db 1 MKLLKVAIAAIVFSGSAGVVPQYGGGNGHGGGNSGPNSELNIYQYGGNSALALQ 60
QY 61 SDARKSETTITQSGYNGADVGCGADNSTIELTQNGFRNNATIDQWAKNSDITVQYGG 120
Db 61 TDARNSLTITQSGGNGADVGCGSDSSIDLQRFNGSATLDQWNGKNSMTVKQFGG 120
QY 121 NNAALVNOTASDSSVMVROVQFGNNATANQY 151
Db 121 GNGAAMDQTASNSVNVTVQFGNNATAHQY 151

```

## RESULT 2

```

US-09-741-873B-4
; Sequence 4, Application US/09741-873B
; Publication No. US20040096965A9
; GENERAL INFORMATION:
; APPLICANT: Normark, Staffan
; APPLICANT: Olsen, Arne
; TITLE OF INVENTION: Fibronectin Binding Protein As Well As Its Preparation
; FILE REFERENCE: 012889-084
; CURRENT APPLICATION NUMBER: US/09/741,873B
; PRIOR APPLICATION NUMBER: SE 8801723-1
; PRIOR FILING DATE: 1998-05-06
; PRIOR APPLICATION NUMBER: US 08/978,878
; PRIOR FILING DATE: 1997-11-26
; PRIOR APPLICATION NUMBER: US 07/347,189
; PRIOR FILING DATE: 1989-05-04
; PRIOR APPLICATION NUMBER: US 07/789,437
; PRIOR FILING DATE: 1991-11-06
; PRIOR APPLICATION NUMBER: US 07/970,846
; PRIOR FILING DATE: 1992-11-03
; PRIOR APPLICATION NUMBER: US 08/187,865
; PRIOR FILING DATE: 1994-01-28
; PRIOR APPLICATION NUMBER: US 08/318,519
; PRIOR FILING DATE: 1994-10-05
; NUMBER OF SEQ ID NOS: 10
; SOFTWARE: Patentin version 3.0
; SEQ ID NO 4
; LENGTH: 151
; TYPE: PRT
; ORGANISM: Escherichia coli
US-09-741-873B-4

```

```

Query Match 66.7%; Score 520; DB 12; Length 151;
Best Local Similarity 68.2%; Pred. No. 2.2e-45;
Matches 103; Conservative 21; Mismatches 27; Indels 0; Gaps 0;

QY 1 MKLLKVAIAAIVVSGSALAGVVPQYGGGNGHGGGNSGPNSELNIYQYGGNSALALQ 60
Db 1 MKLLKVAIAAIVFSGSAGVVPQYGGGNGHGGGNSGPNSELNIYQYGGNSALALQ 60
QY 61 SDARKSETTITQSGYNGADVGCGADNSTIELTQNGFRNNATIDQWAKNSDITVQYGG 120
Db 61 TDARNSLTITQSGGNGADVGCGSDSSIDLQRFNGSATLDQWNGKNSMTVKQFGG 120
QY 121 NNAALVNOTASDSSVMVROVQFGNNATANQY 151
Db 121 GNGAAMDQTASNSVNVTVQFGNNATAHQY 151

```

## RESULT 3

```

US-09-741-873B-2
; Sequence 2, Application US/09741873B
; Publication No. US20020081722A1
; GENERAL INFORMATION:
; APPLICANT: Normark, Staffan
; APPLICANT: Olsen, Arne
; TITLE OF INVENTION: Fibronectin Binding Protein As Well As Its Preparation
; FILE REFERENCE: 012889-084
; CURRENT APPLICATION NUMBER: US/09/741,873B
; CURRENT FILING DATE: 2003-04-04

```

```

; PRIOR APPLICATION NUMBER: SE 8801723-1
; PRIOR FILING DATE: 1998-05-06
; PRIOR APPLICATION NUMBER: US 08/978,878
; PRIOR FILING DATE: 1997-11-26
; PRIOR APPLICATION NUMBER: US 07/347,189
; PRIOR FILING DATE: 1989-05-04
; PRIOR APPLICATION NUMBER: US 07/789,437
; PRIOR FILING DATE: 1991-11-06
; PRIOR APPLICATION NUMBER: US 07/970,846
; PRIOR FILING DATE: 1992-11-03
; PRIOR APPLICATION NUMBER: US 08/187,865
; PRIOR FILING DATE: 1994-01-28
; PRIOR APPLICATION NUMBER: US 08/318,519
; PRIOR FILING DATE: 1994-10-05
; NUMBER OF SEQ ID NOS: 10
; SOFTWARE: Patentin version 3.0
; SEQ ID NO 2
; LENGTH: 131
; TYPE: PRT
; ORGANISM: Escherichia coli
US-09-741-873B-2

Query Match 56.9%; Score 444; DB 12; Length 131;
Best Local Similarity 65.6%; Pred. No. 1.2e-37;
Matches 86; Conservative 19; Mismatches 26; Indels 0; Gaps 0;

QY 21 GVPPQWGGGNGHGGGNSGPNSELNIYQYGGNSALALQTDARNSLTITQHGNGAD 80
Db 1 GVPPQYGGGNGHGGGNSGPNSELNIYQYGGNSALALQTDARNSLTITQHGNGAD 60
QY 81 VQCADNSTIELTQNGFRNNATIDQWAKNSDITVQYGGNSALALVNOTASDSSVMVROV 140
Db 61 VQCGSDSSIDLQRFNGSATLDQWNGKNSMTVKQFGGNGAAVDQTASNSVNVTVQ 120
QY 141 GFGNNATANQY 151
Db 121 GFGNNATAHQY 131

```

## RESULT 4

```

US-09-741-873B-2
; Sequence 2, Application US/09741873B
; Publication No. US20040096965A9
; GENERAL INFORMATION:
; APPLICANT: Normark, Staffan
; APPLICANT: Olsen, Arne
; TITLE OF INVENTION: Fibronectin Binding Protein As Well As Its Preparation
; FILE REFERENCE: 012889-084
; CURRENT APPLICATION NUMBER: US/09/741,873B
; CURRENT FILING DATE: 2003-04-04
; PRIOR APPLICATION NUMBER: SE 8801723-1
; PRIOR FILING DATE: 1998-05-06
; PRIOR APPLICATION NUMBER: US 08/978,878
; PRIOR FILING DATE: 1997-11-26
; PRIOR APPLICATION NUMBER: US 07/347,189
; PRIOR FILING DATE: 1989-05-04
; PRIOR APPLICATION NUMBER: US 07/789,437
; PRIOR FILING DATE: 1991-11-06
; PRIOR APPLICATION NUMBER: US 07/970,846
; PRIOR FILING DATE: 1992-11-03
; PRIOR APPLICATION NUMBER: US 08/187,865
; PRIOR FILING DATE: 1994-01-28
; PRIOR APPLICATION NUMBER: US 08/318,519
; PRIOR FILING DATE: 1994-10-05
; NUMBER OF SEQ ID NOS: 10
; SOFTWARE: Patentin version 3.0
; SEQ ID NO 2
; LENGTH: 131
; TYPE: PRT
; ORGANISM: Escherichia coli
US-09-741-873B-2

Query Match 56.9%; Score 444; DB 12; Length 131;

```

Best Local Similarity 65.6%; Pred. No. 1.2e-37;  
Matches 86; Conservative 19; Mismatches 26; Indels 0; Gaps 0;  
QY 21 GVPQWGGGNGHGGSSGPDYDQLVTRVVTHEMAHALQSDARKSETTITQSGYNGAD 80  
D 1 GVPVQGGGNGHGGSSGPNSELNTYQGGNSALALQTDARNSLDITFHGGGNGAD 60  
QY 81 VQGGADNSTIELTQNGFRNATIDOWNAKNSDITVQYGGNNAALVNCOTASDSSVMVROV 140  
D 61 VQGGSDSSIDITQNGFRNATIDOWNKNSBMTVQFGGGNGAAVDQTSNNSVNTQV 120  
QY 141 GFNNATANQY 151  
D 121 GFNNATAHQY 131  
RESULT 5  
US-10-369-493-20638  
; Sequence 20638, Application US/10369493  
; Publication No. US20030233675A1  
; GENERAL INFORMATION:  
; APPLICANT: Cao, Yongwei  
; APPLICANT: Hinkle, Gregory J.  
; APPLICANT: Slater, Steven C.  
; APPLICANT: Goldman, Barry S.  
; APPLICANT: Chen, Xianfeng  
; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF  
; FILE REFERENCE: 38-10(52052)B  
; CURRENT APPLICATION NUMBER: US/10/369,493  
; CURRENT FILING DATE: 2003-02-28  
; PRIOR APPLICATION NUMBER: US 60/360,039  
; PRIOR FILING DATE: 2002-02-21  
; NUMBER OF SEQ ID NOS: 47374  
; SEQ ID NO 20638  
; LENGTH: 445  
; TYPE: PRT  
; ORGANISM: Rhodospseudomonas palustris  
; FEATURE:  
; NAME/KEY: unsure  
; LOCATION: (1)..(445)  
; OTHER INFORMATION: unsure at all Xaa locations  
US-10-369-493-20638  
Query Match 13.9%; Score 108.5; DB 15; Length 445;  
Best Local Similarity 27.5%; Pred. No. 0.015;  
Matches 46; Conservative 19; Mismatches 61; Indels 41; Gaps 6;  
QY 7 AAFAA-----IVVSGSALAGVVPWGCGG-----NHNGG-----GNSSGPDYDQLVTRV 50  
D 19 AAFADSNVTYLNQTDGQOANITQSGNGSVGAFNGSGFLQENGLTSGAN---LLT-- 73  
QY 51 VTHEMAHALQSDARKSETTITQSGYNGADVGGADNSTIELTQNGFRNATIDOWNAXN 110  
D 74 ----VKQSGNSNSVGRDIQCKQSGAGNSAAIFQEGTGSDVELQQTCTNSGAVPSGNWTN 129  
QY 111 -----SDITVQYGGNNAALVNCOTASDSSVMVROV 141  
D 130 DPGVFNKITQDSSSSGKSVIQDGKNNVFSIXQGTGNSTSVNQIG 176  
RESULT 6  
US-10-282-122A-44999  
; Sequence 44999, Application US/10282122A  
; Publication No. US20040029129A1  
; GENERAL INFORMATION:  
; APPLICANT: Wang, Liangsu  
; APPLICANT: Zamudio, Carlos  
; APPLICANT: Malone, Cheryl  
; APPLICANT: Haselbeck, Robert  
; APPLICANT: Ohlsen, Kari  
; APPLICANT: Zyskind, Judith  
; APPLICANT: Wall, Daniel

APPLICANT: Trawick, John  
APPLICANT: Carr, Grant  
APPLICANT: Yamamoto, Robert  
APPLICANT: Forsyth, R.  
APPLICANT: Xu, H.  
TITLE OF INVENTION: Identification of Essential Genes in Microorganisms  
FILE REFERENCE: ELITRA.034A  
CURRENT APPLICATION NUMBER: US/10/282,122A  
CURRENT FILING DATE: 2003-02-20  
PRIOR APPLICATION NUMBER: 60/191,078  
PRIOR FILING DATE: 2000-03-21  
PRIOR APPLICATION NUMBER: 60/206,848  
PRIOR FILING DATE: 2000-05-23  
PRIOR APPLICATION NUMBER: 60/207,727  
PRIOR FILING DATE: 2000-05-26  
PRIOR APPLICATION NUMBER: 60/230,335  
PRIOR FILING DATE: 2000-09-06  
PRIOR APPLICATION NUMBER: 60/230,347  
PRIOR FILING DATE: 2000-09-09  
PRIOR APPLICATION NUMBER: 60/242,578  
PRIOR FILING DATE: 2000-10-23  
PRIOR APPLICATION NUMBER: 60/253,625  
PRIOR FILING DATE: 2000-11-27  
PRIOR APPLICATION NUMBER: 60/257,931  
PRIOR FILING DATE: 2000-12-22  
PRIOR APPLICATION NUMBER: 60/267,636  
PRIOR FILING DATE: 2001-02-09  
PRIOR APPLICATION NUMBER: 60/269,308  
PRIOR FILING DATE: 2001-02-16  
Remaining Prior Application data removed - See File Wrapper or PALM.  
NUMBER OF SEQ ID NOS: 78614  
SOFTWARE: PatentIn version 3.1  
SEQ ID NO 44999  
LENGTH: 974  
TYPE: PRT  
ORGANISM: Acinetobacter baumannii  
US-10-282-122A-44999  
Query Match 11.9%; Score 93; DB 12; Length 974;  
Best Local Similarity 24.5%; Pred. No. 1.5;  
Matches 36; Conservative 23; Mismatches 52; Indels 36; Gaps 7;  
QY 15 SGSALAGVVPWGCGGNGHGG--GNSSGPDYDQLVTRVVTHEM-----AHALQSDA 63  
D 299 AGNGIA-----SGNGEHNYGIGNGGDDVD--ITAPITGVNIFSGNFTLIGNSSSSSV 350  
QY 64 RKSETTITQS-----GYNGADVGGADNSTIELTQNGF-----RNNATIDOWNAKNS 111  
D 351 NTAPTTNTVNDNTIDNGSGGTGSGNGSGDGLLGAASNGGHNHYGIGNGGDDV 410  
QY 112 DIT-----VGOYGGNNAALVNCOTASDS 133  
D 411 DITAPITGVFNFSGNSFLIGNSSSS 437  
RESULT 7  
US-10-647-057-4  
; Sequence 4, Application US/10647057  
; Publication No. US20040047871A1  
; GENERAL INFORMATION:  
; APPLICANT: KANSAS STATE UNIVERSITY RESEARCH FOUNDATION  
; APPLICANT: NAGARAJA, T.  
; APPLICANT: STEWART, GEORGE  
; APPLICANT: NARAYANAN, SANJEEV  
; APPLICANT: CHENGAPPA, M.  
; TITLE OF INVENTION: RECOMBINANT FUSOBACTERIUM NECROPHORUM LEUKOTOXIN VACCINE AND PREP  
; FILE REFERENCE: PCT-30962  
; CURRENT APPLICATION NUMBER: US/10/647,057  
; CURRENT FILING DATE: 2003-08-22  
; NUMBER OF SEQ ID NOS: 15  
; SOFTWARE: PatentIn version 3.0  
; SEQ ID NO 4

```

; LENGTH: 580
; TYPE: PRT
; ORGANISM: Fusobacterium necrophorum
; US-10-647-057-4

Query Match      11.7%; Score 91.5; DB 12; Length 580;
Best Local Similarity 23.1%; Pred. No. 1.1;
Matches 40; Conservative 18; Mismatches 56; Indels 59; Gaps 6;

QY 5 KVAFAAIVVSGSALAGVVPQWGGGNNH-----GGNSSGPDYDQLVTRVVTHEMAH 57
Db 257 KKAKEGAVVNAASVAGTOKSAGGVAAVTVKKNKASGSNKAGD-----KH 302
QY 58 ALQSDARKSETTITQ-----SGYNGA-----DVGGADNSTIELTQNGFRN 99
Db 303 AKHVNVAKSVTVVNAASGASKDASGSGGAWDSNDTAKVDKGRSADSNV----- 352
QY 100 NATIDQWNAKNSDITYQYGGNNAL---VNCTASDSSVMVRQVGFNNATAN 149
Db 353 -----NANNSGVNVAGTAGSSTAVGAAANNTHNKTSATGTVKNSGKNTKVN 398

RESULT 8
US-09-793-306-146
; Sequence 146, Application US/09793306
; Patent No. US20020098200A1
; GENERAL INFORMATION:
; APPLICANT: Campos-Neto, Antonio
; APPLICANT: Skeiky, Yasir
; APPLICANT: Owendale, Pamela
; APPLICANT: Jen, Shylian
; APPLICANT: Lodes, Michael
; APPLICANT: Corixa Corporation
; TITLE OF INVENTION: Compounds and Methods for Diagnosis and Immunotherapy
; TITLE OF INVENTION: of Tuberculosis
; FILE REFERENCE: 014058-00874035
; CURRENT APPLICATION NUMBER: US/09/793,306
; CURRENT FILING DATE: 2001-02-26
; PRIOR APPLICATION NUMBER: US 60/185,037
; PRIOR FILING DATE: 2000-02-25
; PRIOR APPLICATION NUMBER: US 60/223,828
; PRIOR FILING DATE: 2000-08-08
; NUMBER OF SEQ ID NOS: 164
; SOFTWARE: Patentin Ver. 2.1
; SEQ ID NO 146
; LENGTH: 597
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence:mTTC#3-His
; US-09-793-306-146

Query Match      11.7%; Score 91; DB 9; Length 597;
Best Local Similarity 27.8%; Pred. No. 1.3;
Matches 35; Conservative 13; Mismatches 48; Indels 30; Gaps 5;

QY 26 WGGGNNHGGNSSGPDYDQLVTRVVTHEMAHALQSDARKSETTITQSGYNGADVGOGA 85
Db 358 FGNSGNNNIGFNSG-----NNNVGFFNSGNNFNGAGDINTGF 398
QY 86 DNSTIELTQNGFRNATIDQ---WNKNSDITYQYGGNNALVNCTASDSSVMVRQVGF 143
Db 399 GNAGD---TWTGCGAGFFNNGIGNAGNEDMGVNGSFFNVGVGN---AGNQS-----VGFG 449
QY 144 NNATAN 149
Db 450 NAGTLN 455

RESULT 9
US-10-233-449-7
; Sequence 7, Application US/10233449
; Publication No. US20030211040A1

```

```

; GENERAL INFORMATION:
; APPLICANT: Greengard, Paul
; APPLICANT: Repaske, David
; APPLICANT: Snyder, Gretchen
; TITLE OF INVENTION: PHOSPHODIESTERASE ACTIVITY AND REGULATION OF
; TITLE OF INVENTION: PHOSPHODIESTERASE
; TITLE OF INVENTION: 1B-MEDIATED SIGNALING IN BRAIN
; FILE REFERENCE: 11181-010-999
; CURRENT APPLICATION NUMBER: US/10/233,449
; CURRENT FILING DATE: 2002-09-03
; PRIOR APPLICATION NUMBER: 60/316,320
; PRIOR FILING DATE: 2001-08-30
; NUMBER OF SEQ ID NOS: 7
; SOFTWARE: Patentin version 3.0
; SEQ ID NO 7
; LENGTH: 906
; TYPE: PRT
; ORGANISM: Homo sapiens
; US-10-233-449-7

Query Match      11.5%; Score 90; DB 12; Length 906;
Best Local Similarity 22.6%; Pred. No. 2.8;
Matches 45; Conservative 26; Mismatches 56; Indels 72; Gaps 10;

QY 6 VAAFAAIVVSGSALAGVVPQWGGGNNHGGNSSGPDYDQL-----VTRVVTHE----M 55
Db 250 VTGFQLVNYTDTIPAKIMQW-----KNSDARDHTRDVWKRPKYTSALTYDGVKYM 300
QY 56 AHALQSDARKSETTITQSGYNGADV-----GOGAD-----NSTIELTQ 94
Db 301 ABAFQSLRRQ---RIDLSRRGNAGDCCLANPAVWPWGCGIDIQRALQVRFEGLTGVQFNE 357
QY 95 NGFNNAT-----IDQWNAKNSDI---TVGYGGNNAALVNQT-----ASD 132
Db 358 KGRRTNVTLHVIEKMHDSIRKIGVWNEDEKFPVPAATDAQAGDSSVQNRVTIVTTILED 417
QY 133 SSVVRQVGFNNATANQY 151
Db 418 PYVMLKK-----NANQF 429

RESULT 10
US-10-251-661-2
; Sequence 2, Application US/10251661
; Publication No. US2003016655A1
; GENERAL INFORMATION:
; APPLICANT: Alberini, Cristina M.
; APPLICANT: Bear, Mark F.
; TITLE OF INVENTION: Methods and Compositions for Regulating
; TITLE OF INVENTION: Memory Consolidation
; FILE REFERENCE: 3499.1001-003
; CURRENT APPLICATION NUMBER: US/10/251,661
; CURRENT FILING DATE: 2002-09-20
; PRIOR APPLICATION NUMBER: 60/193,614
; PRIOR FILING DATE: 2000-03-31
; PRIOR APPLICATION NUMBER: PCT/US01/10661
; PRIOR FILING DATE: 2001-04-02
; NUMBER OF SEQ ID NOS: 12
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 2
; LENGTH: 906
; TYPE: PRT
; ORGANISM: Homo sapiens
; US-10-251-661-2

Query Match      11.5%; Score 90; DB 14; Length 906;
Best Local Similarity 22.8%; Pred. No. 2.8;
Matches 45; Conservative 26; Mismatches 56; Indels 72; Gaps 10;

QY 6 VAAFAAIVVSGSALAGVVPQWGGGNNHGGNSSGPDYDQL-----VTRVVTHE----M 55
Db 250 VTGFQLVNYTDTIPAKIMQW-----KNSDARDHTRDVWKRPKYTSALTYDGVKYM 300

```

QY 56 AHALQSDARKSETTITQSGYNGADV-----GOGAD-----NSTIELTQ 94  
DB 301 AEAFQSLERQ---RIDISRGNGDCLANPAVPGGQIDIQALQOQVPEGLTGNVQFNE 357  
QY 95 NGFRNNAT-----IDOWNAKNSDI---TVQYGGNNAALVNOT-----ASD 132  
DB 358 KGRNTYTLHVIEKMDGIRKIGYWNEDDKFVPAATDAQAGGNSVQNRITYIVTILED 417  
QY 133 SSVVMVQVFGNNGNATQY 151  
DB 418 PYVWLKK-----NANQF 429

RESULT 11  
US-10-408-765A-998  
; Sequence 998, Application US/10408765A  
; Publication No. US20040101874A1  
; GENERAL INFORMATION:  
; APPLICANT: Ghosh, Soumitra S.  
; APPLICANT: Fahy, Eoin D.  
; APPLICANT: Zhang, Bing  
; APPLICANT: Gibson, Bradford W.  
; APPLICANT: Taylor, Steven W.  
; APPLICANT: Glenn, Gary M.  
; APPLICANT: Warnock, Dale E.  
; TITLE OF INVENTION: TARGETS FOR THERAPEUTIC INTERVENTION  
; TITLE OF INVENTION: IDENTIFIED IN THE MITOCHONDRIAL PROTEOME  
; FILE REFERENCE: 66088.465  
; CURRENT APPLICATION NUMBER: US/10/408,765A  
; CURRENT FILING DATE: 2003-04-04  
; NUMBER OF SEQ ID NOS: 3077  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 998  
; LENGTH: 1448  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-10-408-765A-998

Query Match 11.4%; Score 89; DB 16; Length 1448;  
Best Local Similarity 24.4%; Pred. No. 6.5;  
Matches 39; Conservative 18; Mismatches 47; Indels 56; Gaps 7;  
QY 10 AAVVSGSALAGVV-----PWG-----GGN-----H 32  
DB 307 AAAAKSGHAWSGAANKEDKPTWGPPEPKSQHWGDQSRNPASWAGGSDWADSSVLGH 366  
QY 33 NGGNSGSDPDYDQVTRVVTHEMAHALQSDARKSETTIT-QSGYNGGADVQGGADNSTIE 91  
DB 367 LGDGKONGSGWD-----ADSNRSGSGWNDITRSGNSGWSGKSTNTKANPGTNWGE 415  
QY 92 LTQGFRRNATIDOWNAKNSDITVQYGGNNAALVNOTAS 131  
DB 416 TLKPGPOON-----WASKPDQNNVSNWGG--AASVXQTGT 448

RESULT 12  
US-10-424-599-203972  
; Sequence 203972, Application US/10424599  
; Publication No. US20040031072A1  
; GENERAL INFORMATION:  
; APPLICANT: La Rosa, Thomas J  
; APPLICANT: Kovalic, David K  
; APPLICANT: Zhou, Yihua  
; APPLICANT: Cao, Yongwei  
; TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated With  
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement  
; FILE REFERENCE: 38-21(53223)B  
; CURRENT APPLICATION NUMBER: US/10/424,599  
; CURRENT FILING DATE: 2003-04-28  
; NUMBER OF SEQ ID NOS: 285684  
; SEQ ID NO 203972  
; LENGTH: 204  
; TYPE: PRT

; ORGANISM: Glycine max  
; FEATURE:  
; OTHER INFORMATION: Clone ID: PAT\_MRT3847\_26213C.1.1.pap  
US-10-424-599-203972

Query Match 11.3%; Score 88.5; DB 12; Length 204;  
Best Local Similarity 25.7%; Pred. No. 0.61;  
Matches 35; Conservative 10; Mismatches 30; Indels 61; Gaps 7;  
QY 21 GVVPQWGGGNGHGGNSGSDPDYDQVTRVVTHEMAHALQSDARKSETTITQSGYNGAD 80  
DB 106 GRVP-W-GGSHERGYSYG-----NSD-----SGWQGGD 134  
QY 81 VQGGADNSTIELTQNGFRNNATIDOWNAKNSDIT-----VQYGGNNAALVNOTAS 131  
DB 135 QGRGSDG-----NGQGRGWRGNGNSNEERKNDEESRGSWGSGFNGAGSGNENS- 182  
QY 132 DSSVMVQVFGNNGNAT 147  
DB 183 -----GWGKNAT 189

RESULT 13  
US-10-425-114-37715  
; Sequence 37715, Application US/10425114  
; Publication No. US20040034888A1  
; GENERAL INFORMATION:  
; APPLICANT: Liu, Jingdong  
; APPLICANT: Zhou, Yihua  
; APPLICANT: Kovalic, David K.  
; APPLICANT: Screen, Steven E.  
; APPLICANT: Tabaska, Jack E.  
; APPLICANT: Cao, Yongwei  
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With  
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement  
; FILE REFERENCE: 38-21(53313)B  
; CURRENT APPLICATION NUMBER: US/10/425,114  
; CURRENT FILING DATE: 2003-04-28  
; NUMBER OF SEQ ID NOS: 73128  
; SEQ ID NO 37715  
; LENGTH: 244  
; TYPE: PRT  
; ORGANISM: Glycine max  
; FEATURE:  
; OTHER INFORMATION: Clone ID: LIB3170-053-G12\_FLI.1.pap  
US-10-425-114-37715

Query Match 11.3%; Score 88.5; DB 12; Length 244;  
Best Local Similarity 22.9%; Pred. No. 0.76;  
Matches 32; Conservative 21; Mismatches 48; Indels 39; Gaps 6;  
QY 27 GGGGNGHGGNSGSDPDY-----DQVTRVVTHEMAHALQSDARKSETTITQSG- 74  
DB 23 GGGGNNNNNNNEGTNYFATSSAPLSFNSVNASALSHMSATALQKAAQMGATTSSNGGT 82  
QY 75 -----YNGGADVQGGADNSTIELTQNGFRNNATIDOWNAKNSDITVQYGGNNAALVN- 127  
DB 83 ASLLKSPGSSASSSGGGGSKLV-----NAAN--YVSGMFGGNH--VNE 122  
QY 128 QTASDSSVMVQVFGNNGNAT 147  
DB 123 QSNSNLQDLNMSFAVGNGSS 142

RESULT 14  
US-10-437-963-114193  
; Sequence 114193, Application US/10437963  
; Publication No. US20040123343A1  
; GENERAL INFORMATION:  
; APPLICANT: La Rosa, Thomas J.  
; APPLICANT: Kovalic, David K.  
; APPLICANT: Zhou, Yihua  
; APPLICANT: Cao, Yongwei

Search completed: August 2, 2004, 15:36:11  
Job time : 37.8 secs

APPLICANT: Wu, Wei  
APPLICANT: Boukharov, Andrey A.  
APPLICANT: Barbazuk, Brad  
APPLICANT: Li, Ping  
TITLE OF INVENTION: Rice Nucleic Acid Molecules and Other Molecules Associated With  
TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement  
FILE REFERENCE: 38-21(53221)B  
CURRENT APPLICATION NUMBER: US/10/437,963  
CURRENT FILING DATE: 2003-05-14  
NUMBER OF SEQ ID NOS: 204966  
SEQ ID NO 114193  
LENGTH: 253  
TYPE: PRT  
ORGANISM: Oryza sativa  
FEATURE:  
OTHER INFORMATION: Clone ID: PAT\_MRT4530\_17908C.1.pap  
US-10-437-963-114193

Query Match 11.3%; Score 88.5; DB 16; Length 253;  
Best Local Similarity 26.6%; Pred. No. 0.8;  
Matches 38; Conservative 14; Mismatches 48; Indels 43; Gaps 6;  
QY 4 LKVA--FRAIVVSGSALAGVFPWG-----GGGNHNGG-----NSSGP 41  
Db 65 IKLAALGFVLLSIGLSAARVERYSSESGGTNGEGGGYNGGGVGGSGGAGSGSGG 124  
QY 42 DYDLVTRVVTHEMAHALOSDARKSETTITQSGY-----GNGADVGGGADNSTIELTQNGF 97  
Db 125 NYC-----AHASGGGGG-----GGYSQYGGSGSGSGSGSSEYTONGG 167  
QY 98 RNNATIDOWNAKNSDITVGQYGG 120  
Db 168 YGAGSSSSAGSGGAGAGGAGG 190

RESULT 15  
US-10-424-599-232271  
Sequence 232271, Application US/10424599  
Publication No. US20040031072A1  
GENERAL INFORMATION:  
APPLICANT: La Rosa Thomas J  
APPLICANT: Zhou Yihua  
APPLICANT: Kovalic David K  
TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated With  
TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement  
FILE REFERENCE: 38-21(53223)B  
CURRENT APPLICATION NUMBER: US/10/424,599  
CURRENT FILING DATE: 2003-04-28  
NUMBER OF SEQ ID NOS: 285684  
SEQ ID NO 232271  
LENGTH: 283  
TYPE: PRT  
ORGANISM: Glycine max  
FEATURE:  
OTHER INFORMATION: Clone ID: PAT\_MRT3847\_51764C.1.pap  
US-10-424-599-232271

Query Match 11.3%; Score 88.5; DB 12; Length 283;  
Best Local Similarity 22.9%; Pred. No. 0.92;  
Matches 32; Conservative 21; Mismatches 48; Indels 39; Gaps 6;  
QY 27 GGGGNHNGGNSGPDY-----DOLVTRVVTHEMAHALOSDARKSETTITQSG- 74  
Db 62 GGGGNHNNNNNNEGTYFATSGAPSLFNSVNASALSHVSATALLQKAAQMGATTSNGGT 121  
QY 75 -----YNGADVGGGADNSTIELTQNGPRNNATIDOWNAKNSDITVGQYGNNAALVN- 127  
Db 122 ASLLKSPGSGSSSSGGGSKLV-----NAAN--YVSGVFGGNH--VNE 161  
QY 128 QTASDSSVNVROVFGNNAT 147  
Db 162 QSNLNQDLNLSFAVGGNSS 181



GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

QM protein - protein search, using sw model

Run on: August 2, 2004, 14:48:33 ; Search time 167.9 seconds  
(without alignments)  
877.809 Million cell updates/sec

Title: US-09-543-407-18  
Perfect score: 780  
Sequence: 1 MKLLKVAFAAIVSGSALA.....DSSVMVROVGFNNATANYQ 151

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 6019581 seqs, 976053577 residues

Total number of hits satisfying chosen parameters: 6019581

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Pending Patents\_AA\_Main:\*  
1: /cgn2\_6/prodata/2/paa/pctus COMB.pep.\*  
2: /cgn2\_6/prodata/2/paa/us06 COMB.pep.\*  
3: /cgn2\_6/prodata/2/paa/us07 COMB.pep.\*  
4: /cgn2\_6/prodata/2/paa/us08 COMB.pep.\*  
5: /cgn2\_6/prodata/2/paa/us081 COMB.pep.\*  
6: /cgn2\_6/prodata/2/paa/us082 COMB.pep.\*  
7: /cgn2\_6/prodata/2/paa/us083 COMB.pep.\*  
8: /cgn2\_6/prodata/2/paa/us084 COMB.pep.\*  
9: /cgn2\_6/prodata/2/paa/us085 COMB.pep.\*  
10: /cgn2\_6/prodata/2/paa/us086 COMB.pep.\*  
11: /cgn2\_6/prodata/2/paa/us087 COMB.pep.\*  
12: /cgn2\_6/prodata/2/paa/us088 COMB.pep.\*  
13: /cgn2\_6/prodata/2/paa/us089 COMB.pep.\*  
14: /cgn2\_6/prodata/2/paa/us090 COMB.pep.\*  
15: /cgn2\_6/prodata/2/paa/us091 COMB.pep.\*  
16: /cgn2\_6/prodata/2/paa/us092 COMB.pep.\*  
17: /cgn2\_6/prodata/2/paa/us093 COMB.pep.\*  
18: /cgn2\_6/prodata/2/paa/us094 COMB.pep.\*  
19: /cgn2\_6/prodata/2/paa/us095 COMB.pep.\*  
20: /cgn2\_6/prodata/2/paa/us096 COMB.pep.\*  
21: /cgn2\_6/prodata/2/paa/us097A COMB.pep.\*  
22: /cgn2\_6/prodata/2/paa/us097B COMB.pep.\*  
23: /cgn2\_6/prodata/2/paa/us098 COMB.pep.\*  
24: /cgn2\_6/prodata/2/paa/us099A COMB.pep.\*  
25: /cgn2\_6/prodata/2/paa/us099B COMB.pep.\*  
26: /cgn2\_6/prodata/2/paa/us100 COMB.pep.\*  
27: /cgn2\_6/prodata/2/paa/us101 COMB.pep.\*  
28: /cgn2\_6/prodata/2/paa/us102 COMB.pep.\*  
29: /cgn2\_6/prodata/2/paa/us103 COMB.pep.\*  
30: /cgn2\_6/prodata/2/paa/us104 COMB.pep.\*  
31: /cgn2\_6/prodata/2/paa/us106 COMB.pep.\*  
32: /cgn2\_6/prodata/2/paa/us107 COMB.pep.\*  
33: /cgn2\_6/prodata/2/paa/us60 COMB.pep.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
------------	-------	-------------	--------	----	-------------

1	780	100.0	151	19	US-09-543-407-18	Sequence 18, Appl
2	691	88.6	151	19	US-09-543-407-5	Sequence 5, Appl
3	686	87.5	151	6	US-08-233-642A-57	Sequence 57, Appl
4	675	86.9	151	19	US-09-543-407-26	Sequence 26, Appl
5	622	79.7	151	19	US-09-543-407-20	Sequence 20, Appl
6	618	79.2	151	19	US-09-543-407-16	Sequence 16, Appl
7	613	78.6	151	19	US-09-543-407-12	Sequence 12, Appl
8	611	77.3	151	19	US-09-543-407-14	Sequence 14, Appl
9	608	77.9	151	19	US-09-543-407-24	Sequence 24, Appl
10	604	77.4	131	19	US-09-543-407-31	Sequence 31, Appl
11	601	77.1	151	19	US-09-543-407-28	Sequence 28, Appl
12	600	76.9	151	19	US-09-543-407-22	Sequence 22, Appl
13	599	76.8	151	19	US-09-543-407-30	Sequence 30, Appl
14	523	67.1	151	19	US-09-543-407-7	Sequence 7, Appl
15	520	66.7	151	13	US-08-978-878-4	Sequence 4, Appl
16	520	66.7	151	21	US-09-741-873B-4	Sequence 2, Appl
17	518	66.4	151	33	US-60-352-946-2	Sequence 2, Appl
18	518	66.4	151	33	US-60-444-371-2	Sequence 2, Appl
19	506	64.9	120	6	US-08-233-642A-55	Sequence 55, Appl
20	483	61.9	109	19	US-09-543-407-34	Sequence 34, Appl
21	457	58.6	158	16	US-09-252-691-5834	Sequence 5834, Ap
22	457	58.6	158	16	US-09-252-691C-5834	Sequence 5834, Ap
23	457	58.6	158	30	US-10-417-886-5834	Sequence 2, Appl
24	444	56.9	131	13	US-08-978-878-2	Sequence 2, Appl
25	444	56.9	131	21	US-09-741-873B-2	Sequence 35, Appl
26	353.5	46.1	109	19	US-09-543-407-35	Sequence 37, Appl
27	276	35.4	68	19	US-09-543-407-37	Sequence 32, Appl
28	250.5	32.1	70	19	US-09-543-407-32	Sequence 39, Appl
29	175	22.4	48	19	US-09-543-407-39	Sequence 39, Appl
30	116	14.9	186	16	US-09-252-691-5833	Sequence 5833, Ap
31	116	14.9	186	16	US-09-252-691C-5833	Sequence 5833, Ap
32	116	14.9	186	30	US-10-417-886-5833	Sequence 5833, Ap
33	108.5	13.9	445	29	US-10-369-493-20638	Sequence 20638, A
34	108.5	13.9	445	33	US-60-360-039-20638	Sequence 20638, A
35	107.5	13.8	151	19	US-09-543-407-8	Sequence 8, Appl
36	106.5	13.7	151	19	US-09-543-407-6	Sequence 6, Appl
37	101	12.9	257	33	US-60-173-464-21553	Sequence 21553, A
38	101	12.9	262	20	US-09-614-150-25818	Sequence 25818, A
39	101	12.9	262	20	US-09-614-150A-25818	Sequence 25818, A
40	101	12.9	262	33	US-60-191-637-25957	Sequence 25957, A
41	101	12.9	262	33	US-60-191-681-20566	Sequence 20566, A
42	100	12.8	1308	27	US-10-179-131-5148	Sequence 5148, Ap
43	99.5	12.8	1249	30	US-10-455-719-358	Sequence 358, App
44	99.5	12.8	1249	33	US-60-385-568-357	Sequence 357, App
45	99.5	12.8	1249	33	US-60-446-775-358	Sequence 358, App

ALIGNMENTS

RESULT 1

US-09-543-407-18  
; Sequence 18, Application US/09543407  
; GENERAL INFORMATION:  
; APPLICANT: White, Aaron P.  
; APPLICANT: Doran, James L.  
; APPLICANT: Collinson, S. Karen  
; APPLICANT: Kay, William W.  
; TITLE OF INVENTION: BACTERIAL FIMBRIAL SYSTEM FOR PRESENTATION OF HETEROLOGOUS PEPTIDE SEQUENCES  
; FILE REFERENCE: 920043.406  
; CURRENT APPLICATION NUMBER: US/09/543,407  
; CURRENT FILING DATE: 2000-04-05  
; NUMBER OF SEQ ID NOS: 59  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 18  
; TYPE: PRT  
; LENGTH: 151  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: Recombinant Salmonella enteritidis 3b afga  
; OTHER INFORMATION: sequence containing the replacement fragment  
; OTHER INFORMATION: encoding PT3 from GP63 of Leishmania major.

US-09-543-407-18

Query Match 100.0%; Score 780; DB 19; Length 151;  
Best Local Similarity 100.0%; Pred. No. 3.7e-75;  
Matches 151; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MKLLKVAFAAIVVSGSALAGVVPQGGGNGHNGGSGPDYDQLVTRVVTHEMAHALQ 60  
DB 1 MKLLKVAFAAIVVSGSALAGVVPQGGGNGHNGGSGPDYDQLVTRVVTHEMAHALQ 60  
QY 61 SDARKSETTITQSGYNGADVGQADNSTIELTQNGFRNNATIDQNAKNSDITVGYGG 120  
DB 61 SDARKSETTITQSGYNGADVGQADNSTIELTQNGFRNNATIDQNAKNSDITVGYGG 120

QY 121 NNAALVNQTASDSSVMVRQVGFNNATANOY 151  
DB 121 NNAALVNQTASDSSVMVRQVGFNNATANOY 151

## RESULT 2

US-09-543-407-5

; Sequence 5, Application US/09543407

; GENERAL INFORMATION:

; APPLICANT: White, Aaron P.

; APPLICANT: Doran, James L.

; APPLICANT: Collinson, S. Karen

; APPLICANT: Kay, William W.

; TITLE OF INVENTION: BACTERIAL FIMBRIAL SYSTEM FOR

; TITLE OF INVENTION: PRESENTATION OF HETEROLOGOUS PEPTIDE SEQUENCES

; FILE REFERENCES: 920043.406

; CURRENT APPLICATION NUMBER: US/09/543.407

; CURRENT FILING DATE: 2000-04-05

; NUMBER OF SEQ ID NOS: 59

; SOFTWARE: FastSeq for Windows Version 4.0

; SEQ ID NO 5

; LENGTH: 151

; TYPE: PRT

; ORGANISM: Salmonella enteritidis

US-09-543-407-5

Query Match 88.6%; Score 691; DB 19; Length 151;  
Best Local Similarity 90.1%; Pred. No. 1.4e-65;  
Matches 136; Conservative 3; Mismatches 11; Indels 0; Gaps 0;

QY 1 MKLLKVAFAAIVVSGSALAGVVPQGGGNGHNGGSGPDYDQLVTRVVTHEMAHALQ 60  
DB 1 MKLLKVAFAAIVVSGSALAGVVPQGGGNGHNGGSGPDYDQLVTRVVTHEMAHALQ 60

QY 61 SDARKSETTITQSGYNGADVGQADNSTIELTQNGFRNNATIDQNAKNSDITVGYGG 120  
DB 61 SDARKSETTITQSGYNGADVGQADNSTIELTQNGFRNNATIDQNAKNSDITVGYGG 120

QY 121 NNAALVNQTASDSSVMVRQVGFNNATANOY 151  
DB 121 NNAALVNQTASDSSVMVRQVGFNNATANOY 151

## RESULT 3

US-08-233-642A-57

; Sequence 57, Application US/08233642A

; GENERAL INFORMATION:

; APPLICANT: Kay, William W.

; APPLICANT: Collinson, S. Karen

; APPLICANT: Clouthier, Sharon C.

; APPLICANT: Doran, James L.

; TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR SALMONELLA-

; TITLE OF INVENTION: BASED VACCINES

; NUMBER OF SEQUENCES: 58

; NUMBER OF SEQUENCES: 58

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Seed and Berry

; STREET: 6300 Columbia Center, 701 Fifth Avenue

; CITY: Seattle

STATE: Washington  
COUNTRY: U.S.A.  
ZIP: 98104-7092  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/233.642A  
FILING DATE: 26-APR-1994  
CLASSIFICATION: 424  
ATTORNEY/AGENT INFORMATION:  
NAME: King, Joshua  
REGISTRATION NUMBER: 35,570  
REFERENCE/DOCKET NUMBER: 920043.403C3  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (206) 622-4900  
TELEFAX: (206) 682-6031  
TELEX: 3723836 SEEDANBERRY  
INFORMATION FOR SEQ ID NO. 57:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 151 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-233-642A-57

Query Match 87.9%; Score 686; DB 6; Length 151;

Best Local Similarity 90.1%; Pred. No. 4.8e-65;

Matches 136; Conservative 3; Mismatches 12; Indels 0; Gaps 0;

QY 1 MKLLKVAFAAIVVSGSALAGVVPQGGGNGHNGGSGPDYDQLVTRVVTHEMAHALQ 60  
DB 1 MKLLKVAFAAIVVSGSALAGVVPQGGGNGHNGGSGPDYDQLVTRVVTHEMAHALQ 60

QY 61 SDARKSETTITQSGYNGADVGQADNSTIELTQNGFRNNATIDQNAKNSDITVGYGG 120  
DB 61 SDARKSETTITQSGYNGADVGQADNSTIELTQNGFRNNATIDQNAKNSDITVGYGG 120

QY 121 NNAALVNQTASDSSVMVRQVGFNNATANOY 151

DB 121 NNAALVNQTASDSSVMVRQVGFNNATANOY 151

## RESULT 4

US-09-543-407-26

; Sequence 26, Application US/09543407

; GENERAL INFORMATION:

; APPLICANT: White, Aaron P.

; APPLICANT: Doran, James L.

; APPLICANT: Collinson, S. Karen

; APPLICANT: Kay, William W.

; TITLE OF INVENTION: BACTERIAL FIMBRIAL SYSTEM FOR

; TITLE OF INVENTION: PRESENTATION OF HETEROLOGOUS PEPTIDE SEQUENCES

; FILE REFERENCES: 920043.406

; CURRENT APPLICATION NUMBER: US/09/543.407

; CURRENT FILING DATE: 2000-04-05

; NUMBER OF SEQ ID NOS: 59

; SOFTWARE: FastSeq for Windows Version 4.0

; SEQ ID NO 26

; LENGTH: 151

; TYPE: PRT

; ORGANISM: Artificial Sequence

; FEATURE:

; OTHER INFORMATION: Recombinant Salmonella enteritidis 3b afga

; OTHER INFORMATION: sequence containing the replacement fragment

; OTHER INFORMATION: encoding P73 from GP63 of Leishmania major.

US-09-543-407-26

Query Match 86.5%; Score 675; DB 19; Length 151;

Best Local Similarity 81.9%; Pred. No. 7.3e-64;

Matches 136; Conservative 0; Mismatches 30; Indels 0; Gaps 2;

```
QY 1 MKLLKVAFAAIVVSGSALAGVVPQWGGGNGHNGGNSGPD-----YDQ 45
Db 1 MKLLKVAFAAIVVSGSALAGVVPQWGGGNGHNGGNSGPDSTLSIYQVGSANAALYDQ 60
QY 46 LVTRVVTTHMAHALQSDARKSETTITQSGYNGADVGQADNSTIELTQNGFRNNATIDQ 105
Db 61 LVTRVVTTHMAHA-----GYNGADVGQADNSTIELTQNGFRNNATIDQ 105
QY 106 WNAXNSDITVQYGGNNAALVNQTSASSVVMVRQVFGNNATANQY 151
Db 106 WNAXNSDITVQYGGNNAALVNQTSASSVVMVRQVFGNNATANQY 151

RESULT 5
US-09-543-407-20
; Sequence 20, Application US/09543407
; GENERAL INFORMATION:
; APPLICANT: White, Aaron P.
; APPLICANT: Doran, James L.
; APPLICANT: Collinson, S. Karen
; APPLICANT: Kay, William W.
; TITLE OF INVENTION: BACTERIAL FIMBRIAL SYSTEM FOR
; PRESENTATION OF HETEROLOGOUS PEPTIDE SEQUENCES
; FILE REFERENCE: 920043.406
; CURRENT APPLICATION NUMBER: US/09/543,407
; CURRENT FILING DATE: 2000-04-05
; NUMBER OF SEQ ID NOS: 59
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 20
; LENGTH: 151
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Recombinant Salmonella enteritidis 3b afga
; OTHER INFORMATION: sequence containing the replacement fragment
; OTHER INFORMATION: encoding PT3 from Gp63 of Leishmania major.
US-09-543-407-20

Query Match 79.7%; Score 622; DB 19; Length 151;
Best Local Similarity 73.6%; Pred. No. 3.6e-58;
Matches 128; Conservative 0; Mismatches 0; Indels 46; Gaps 2;

QY 1 MKLLKVAFAAIVVSGSALAGVVPQWGGGNGHNGGNSGPD-----42
Db 1 MKLLKVAFAAIVVSGSALAGVVPQWGGGNGHNGGNSGPDSTLSIYQVGSANAALYDQ 60
QY 43 -----YDQVTRVVTTHMAHALQSDARKSETTITQSGYNGADVGQADNSTIELTQNGF 97
Db 61 SDARKYDQVTRVVTTHMAHA-----GQADNSTIELTQNGF 97
QY 98 RNNATIDQWNAKNSDITVQYGGNNAALVNQTSASSVVMVRQVFGNNATANQY 151
Db 98 RNNATIDQWNAKNSDITVQYGGNNAALVNQTSASSVVMVRQVFGNNATANQY 151

RESULT 6
US-09-543-407-16
; Sequence 16, Application US/09543407
; GENERAL INFORMATION:
; APPLICANT: White, Aaron P.
; APPLICANT: Doran, James L.
; APPLICANT: Collinson, S. Karen
; APPLICANT: Kay, William W.
; TITLE OF INVENTION: BACTERIAL FIMBRIAL SYSTEM FOR
; PRESENTATION OF HETEROLOGOUS PEPTIDE SEQUENCES
; FILE REFERENCE: 920043.406
; CURRENT APPLICATION NUMBER: US/09/543,407
; CURRENT FILING DATE: 2000-04-05
; NUMBER OF SEQ ID NOS: 59
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 16
; LENGTH: 151
```

```
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Recombinant Salmonella enteritidis 3b afga
; OTHER INFORMATION: sequence containing the replacement fragment
; OTHER INFORMATION: encoding PT3 from Gp63 of Leishmania major.
US-09-543-407-16

Query Match 79.2%; Score 618; DB 19; Length 151;
Best Local Similarity 76.6%; Pred. No. 9.8e-58;
Matches 131; Conservative 0; Mismatches 0; Indels 40; Gaps 2;

QY 1 MKLLKVAFAAIVVSGSALAGVVPQWGGGNGHNGGNSGPDYDQVTRVVTTHMAHA--- 57
Db 1 MKLLKVAFAAIVVSGSALAGV-----YDQVTRVVTTHMAHAG 40
QY 58 -----ALQSDARKSETTITQSGYNGADVGQADNSTIELTQNGFRNN 100
Db 41 PDSTLSIYQVGSANAALQSDARKSETTITQSGYNGADVGQADNSTIELTQNGFRNN 100
QY 101 ATIDQWNAKNSDITVQYGGNNAALVNQTSASSVVMVRQVFGNNATANQY 151
Db 101 ATIDQWNAKNSDITVQYGGNNAALVNQTSASSVVMVRQVFGNNATANQY 151

RESULT 7
US-09-543-407-12
; Sequence 12, Application US/09543407
; GENERAL INFORMATION:
; APPLICANT: White, Aaron P.
; APPLICANT: Doran, James L.
; APPLICANT: Collinson, S. Karen
; APPLICANT: Kay, William W.
; TITLE OF INVENTION: BACTERIAL FIMBRIAL SYSTEM FOR
; PRESENTATION OF HETEROLOGOUS PEPTIDE SEQUENCES
; FILE REFERENCE: 920043.406
; CURRENT APPLICATION NUMBER: US/09/543,407
; CURRENT FILING DATE: 2000-04-05
; NUMBER OF SEQ ID NOS: 59
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 12
; LENGTH: 151
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Recombinant Salmonella enteritidis 3b afga
; OTHER INFORMATION: sequence containing the replacement fragment
; OTHER INFORMATION: encoding PT3 from Gp63 of Leishmania major.
US-09-543-407-12

Query Match 78.6%; Score 613; DB 19; Length 151;
Best Local Similarity 80.8%; Pred. No. 3.4e-57;
Matches 122; Conservative 7; Mismatches 22; Indels 0; Gaps 0;

QY 1 MKLLKVAFAAIVVSGSALAGVVPQWGGGNGHNGGNSGPDYDQVTRVVTTHMAHALQ 60
Db 1 MKLLKVAFAAIVVSGSALAGVVPQWGGGNGHNGGNSGPDSTLSIYQVGSANAALYDQ 60
QY 61 SDARKSETTITQSGYNGADVGQADNSTIELTQNGFRNNATIDQWNAKNSDITVQYGG 120
Db 61 SDARKSETTITQSGYNGADVGQADNSTIELTQNGFRNNATIDQWNAKNSDITVQYGG 120
QY 121 NNAALVNQTSASSVVMVRQVFGNNATANQY 151
Db 121 NNAALVNQVTRVVTTHMAHANNATANQY 151

RESULT 8
US-09-543-407-14
; Sequence 14, Application US/09543407
; GENERAL INFORMATION:
; APPLICANT: White, Aaron P.
; APPLICANT: Doran, James L.
```

APPLICANT: Collinson, S. Karen  
APPLICANT: Kay, William W.  
TITLE OF INVENTION: BACTERIAL FIMBRIAL SYSTEM FOR  
PRESENTATION OF HETEROLOGOUS PEPTIDE SEQUENCES  
FILE REFERENCE: 920043.406  
CURRENT APPLICATION NUMBER: US/09/543,407  
CURRENT FILING DATE: 2000-04-05  
NUMBER OF SEQ ID NOS: 59  
SOFTWARE: FastSeq for Windows Version 4.0  
SEQ ID NO 14  
LENGTH: 151  
TYPE: PRT  
ORGANISM: Artificial Sequence  
FEATURE:  
OTHER INFORMATION: Recombinant Salmonella enteritidis 3b afga  
OTHER INFORMATION: sequence containing the replacement fragment  
OTHER INFORMATION: encoding PT3 from GP63 of Leishmania major.  
US-09-543-407-14

Query Match 78.3%; Score 611; DB 19; Length 151;  
Best Local Similarity 81.5%; Pred. No. 5.5e-57;  
Matches 123; Conservative 6; Mismatches 22; Indels 0; Gaps 0;  
QY 1 MKLLKVAFAAIVVSGSALAGVVPQGGGNGGNSGGPDYDQLVTRVVTHEMAHALQ 60  
DB 1 MKLLKVAFAAIVVSGSALAGVVPQGGGNGGNSGGPDYDQLVTRVVTHEMAHALQ 60  
QY 61 SDARKSETTITQSGYNGADVGCGADNSTIETQNGFRNNATIDQWNAKNSDITVGOYGG 120  
DB 61 SDARKSETTITQSGYNGADVGCGADNSTIETQNGFRNNATIDQWNAKNSDITVGOYDQ 120  
QY 121 NNAALVNQTASDSSVMVRQVGFNNATANQY 151  
DB 121 LVTRVVTHEMAHASVMVRQVGFNNATANQY 151

RESULT 9  
US-09-543-407-24  
Sequence 24, Application US/09543407  
GENERAL INFORMATION:  
APPLICANT: White, Aaron P.  
APPLICANT: Doran, James L.  
APPLICANT: Collinson, S. Karen  
APPLICANT: Kay, William W.  
TITLE OF INVENTION: BACTERIAL FIMBRIAL SYSTEM FOR  
PRESENTATION OF HETEROLOGOUS PEPTIDE SEQUENCES  
FILE REFERENCE: 920043.406  
CURRENT APPLICATION NUMBER: US/09/543,407  
CURRENT FILING DATE: 2000-04-05  
NUMBER OF SEQ ID NOS: 59  
SOFTWARE: FastSeq for Windows Version 4.0  
SEQ ID NO 24  
LENGTH: 151  
TYPE: PRT  
ORGANISM: Artificial Sequence  
FEATURE:  
OTHER INFORMATION: Recombinant Salmonella enteritidis 3b afga  
OTHER INFORMATION: sequence containing the replacement fragment  
OTHER INFORMATION: encoding PT3 from GP63 of Leishmania major.  
US-09-543-407-24

Query Match 77.9%; Score 608; DB 19; Length 151;  
Best Local Similarity 82.1%; Pred. No. 1.2e-56;  
Matches 124; Conservative 5; Mismatches 22; Indels 0; Gaps 0;  
QY 1 MKLLKVAFAAIVVSGSALAGVVPQGGGNGGNSGGPDYDQLVTRVVTHEMAHALQ 60  
DB 1 MKLLKVAFAAIVVSGSALAGVVPQGGGNGGNSGGPDYDQLVTRVVTHEMAHALQ 60  
QY 61 SDARKSETTITQSGYNGADVGCGADNSTIETQNGFRNNATIDQWNAKNSDITVGOYGG 120  
DB 61 SDARKSETTITQSGYNGADVGCGADNSTIETQNGFRNNATIDQWNAKNSDITVGOYDQ 120

QY 121 NNAALVNQTASDSSVMVRQVGFNNATANQY 151  
DB 121 HEMAHANQTASDSSVMVRQVGFNNATANQY 151  
RESULT 10  
US-09-543-407-31  
Sequence 31, Application US/09543407  
GENERAL INFORMATION:  
APPLICANT: White, Aaron P.  
APPLICANT: Doran, James L.  
APPLICANT: Collinson, S. Karen  
APPLICANT: Kay, William W.  
TITLE OF INVENTION: BACTERIAL FIMBRIAL SYSTEM FOR  
PRESENTATION OF HETEROLOGOUS PEPTIDE SEQUENCES  
FILE REFERENCE: 920043.406  
CURRENT APPLICATION NUMBER: US/09/543,407  
CURRENT FILING DATE: 2000-04-05  
NUMBER OF SEQ ID NOS: 59  
SOFTWARE: FastSeq for Windows Version 4.0  
SEQ ID NO 31  
LENGTH: 131  
TYPE: PRT  
ORGANISM: Salmonella enteritidis  
US-09-543-407-31

Query Match 77.4%; Score 604; DB 19; Length 131;  
Best Local Similarity 89.3%; Pred. No. 2.6e-56;  
Matches 117; Conservative 3; Mismatches 11; Indels 0; Gaps 0;  
QY 21 GVVPQGGGNGGNSGGPDYDQLVTRVVTHEMAHALQSDARKSETTITQSGYNGAD 80  
DB 1 GVVPQGGGNGGNSGGPDYDQLVTRVVTHEMAHALQSDARKSETTITQSGYNGAD 80  
QY 81 VGGADNSTIETQNGFRNNATIDQWNAKNSDITVGOYGGNNAALVNQTASDSSVMVRQV 140  
DB 61 VGGADNSTIETQNGFRNNATIDQWNAKNSDITVGOYGGNNAALVNQTASDSSVMVRQV 120  
QY 141 GFGNNATANQY 151  
DB 121 GFGNNATANQY 131

RESULT 11  
US-09-543-407-28  
Sequence 28, Application US/09543407  
GENERAL INFORMATION:  
APPLICANT: White, Aaron P.  
APPLICANT: Doran, James L.  
APPLICANT: Collinson, S. Karen  
APPLICANT: Kay, William W.  
TITLE OF INVENTION: BACTERIAL FIMBRIAL SYSTEM FOR  
PRESENTATION OF HETEROLOGOUS PEPTIDE SEQUENCES  
FILE REFERENCE: 920043.406  
CURRENT APPLICATION NUMBER: US/09/543,407  
CURRENT FILING DATE: 2000-04-05  
NUMBER OF SEQ ID NOS: 59  
SOFTWARE: FastSeq for Windows Version 4.0  
SEQ ID NO 28  
LENGTH: 151  
TYPE: PRT  
ORGANISM: Artificial Sequence  
FEATURE:  
OTHER INFORMATION: Recombinant Salmonella enteritidis 3b afga  
OTHER INFORMATION: sequence containing the replacement fragment  
OTHER INFORMATION: encoding PT3 from GP63 of Leishmania major.  
US-09-543-407-28

Query Match 77.1%; Score 601; DB 19; Length 151;  
Best Local Similarity 80.8%; Pred. No. 6.6e-56;  
Matches 122; Conservative 4; Mismatches 25; Indels 0; Gaps 0;  
QY 1 MKLLKVAFAAIVVSGSALAGVVPQGGGNGGNSGGPDYDQLVTRVVTHEMAHALQ 60

Db 1 M K L L K V A A F A I V V S G S A L A G V V P Q W G G G N H G G S G P D S T L S I Y Q Y G S A N A A L A L Q 60  
QY 61 S D A R K S E T T I T Q S G Y G N G A D V G O G A D N S T I E L T Q N G F R N N A T I D O W N A K N S D I T V G Q Y G G 120  
Db 61 S D A R K S E T T I T Q S G Y G N G A D V G O G A D N S T I E L T Q N G F R N N A T I D O W N A K N S D I T V G Q Y G G 120  
QY 121 N N A A L V N Q T A S D S S V M V R Q V G F G N N A T A N Q Y 151  
Db 121 N N A A L V N Q T A S D S S V M V R Q V G F G N N A T A N Q Y 151

## RESULT 12

US-09-543-407-22  
; Sequence 22, Application US/09543407  
; GENERAL INFORMATION:  
; APPLICANT: White, Aaron P.  
; APPLICANT: Doran, James L.  
; APPLICANT: Collinson, S. Karen  
; APPLICANT: Kay, William W.  
; TITLE OF INVENTION: BACTERIAL FIMBRIAL SYSTEM FOR  
; FILE REFERENCE: 920043.406  
; CURRENT APPLICATION NUMBER: US/09/543,407  
; CURRENT FILING DATE: 2000-04-05  
; NUMBER OF SEQ ID NOS: 59  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 22  
; LENGTH: 151  
; TYPE: PRT  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: Recombinant Salmonella enteritidis 3b afga  
; OTHER INFORMATION: sequence containing the replacement fragment  
; OTHER INFORMATION: encoding PT3 from GP63 of Leishmania major.  
US-09-543-407-22

Query Match 76.9%; Score 600; DB 19; Length 151;  
Best Local Similarity 81.5%; Pred. No. 8.4e-56;  
Matches 123; Conservative 5; Mismatches 23; Indels 0; Gaps 0;  
QY 1 M K L L K V A A F A I V V S G S A L A G V V P Q W G G G N H G G S G P D S T L S I Y Q Y G S A N A A L A L Q 60  
Db 1 M K L L K V A A F A I V V S G S A L A G V V P Q W G G G N H G G S G P D S T L S I Y Q Y G S A N A A L A L Q 60  
QY 61 S D A R K S E T T I T Q S G Y G N G A D V G O G A D N S T I E L T Q N G F R N N A T I D O W N A K N S D I T V G Q Y G G 120  
Db 61 S D A R K S E T T I T Q S G Y G N G A D V G O G A D N S T I E L T Q N G F R N N A T I D O W N A K N S D I T V G Q Y G G 120  
QY 121 N N A A L V N Q T A S D S S V M V R Q V G F G N N A T A N Q Y 151  
Db 121 N N A A L V N Q T A S D S S V M V R Q V G F G N N A T A N Q Y 151

## RESULT 13

US-09-543-407-30  
; Sequence 30, Application US/09543407  
; GENERAL INFORMATION:  
; APPLICANT: White, Aaron P.  
; APPLICANT: Doran, James L.  
; APPLICANT: Collinson, S. Karen  
; APPLICANT: Kay, William W.  
; TITLE OF INVENTION: BACTERIAL FIMBRIAL SYSTEM FOR  
; FILE REFERENCE: 920043.406  
; CURRENT APPLICATION NUMBER: US/09/543,407  
; CURRENT FILING DATE: 2000-04-05  
; NUMBER OF SEQ ID NOS: 59  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 30  
; LENGTH: 151  
; TYPE: PRT  
; ORGANISM: Artificial Sequence

; FEATURE:  
; OTHER INFORMATION: Recombinant Salmonella enteritidis 3b afga  
; OTHER INFORMATION: sequence containing the replacement fragment  
; OTHER INFORMATION: encoding PT3 from GP63 of Leishmania major.  
US-09-543-407-30

Query Match 76.8%; Score 599; DB 19; Length 151;  
Best Local Similarity 81.5%; Pred. No. 1.1e-55;  
Matches 123; Conservative 5; Mismatches 23; Indels 0; Gaps 0;  
QY 1 M K L L K V A A F A I V V S G S A L A G V V P Q W G G G N H G G S G P D S T L S I Y Q Y G S A N A A L A L Q 60  
Db 1 M K L L K V A A F A I V V S G S A L A G V V P Q W G G G N H G G S G P D S T L S I Y Q Y G S A N A A L A L Q 60  
QY 61 S D A R K S E T T I T Q S G Y G N G A D V G O G A D N S T I E L T Q N G F R N N A T I D O W N A K N S D I T V G Q Y G G 120  
Db 61 S D A R K S E T T I T Q S G Y G N G A D V G O G A D N S T I E L T Q N G F R N N A T I D O W N A K N S D I T V G Q Y G G 120  
QY 121 N N A A L V N Q T A S D S S V M V R Q V G F G N N A T A N Q Y 151  
Db 121 N N A A L V N Q T A S D S S V M V R Q V G F G N N A T A N Q Y 151

## RESULT 14

US-09-543-407-7  
; Sequence 7, Application US/09543407  
; GENERAL INFORMATION:  
; APPLICANT: White, Aaron P.  
; APPLICANT: Doran, James L.  
; APPLICANT: Collinson, S. Karen  
; APPLICANT: Kay, William W.  
; TITLE OF INVENTION: BACTERIAL FIMBRIAL SYSTEM FOR  
; FILE REFERENCE: 920043.406  
; CURRENT APPLICATION NUMBER: US/09/543,407  
; CURRENT FILING DATE: 2000-04-05  
; NUMBER OF SEQ ID NOS: 59  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 7  
; LENGTH: 151  
; TYPE: PRT  
; ORGANISM: Escherichia coli  
US-09-543-407-7

Query Match 67.1%; Score 523; DB 19; Length 151;  
Best Local Similarity 68.9%; Pred. No. 1.6e-47;  
Matches 104; Conservative 20; Mismatches 27; Indels 0; Gaps 0;  
QY 1 M K L L K V A A F A I V V S G S A L A G V V P Q W G G G N H G G S G P D S T L S I Y Q Y G S A N A A L A L Q 60  
Db 1 M K L L K V A A F A I V V S G S A L A G V V P Q W G G G N H G G S G P D S T L S I Y Q Y G S A N A A L A L Q 60  
QY 61 S D A R K S E T T I T Q S G Y G N G A D V G O G A D N S T I E L T Q N G F R N N A T I D O W N A K N S D I T V G Q Y G G 120  
Db 61 T D A R N S D L T I T Q H G G G N G A D V G Q G S D S S I D L T Q R G F G N S A T L D Q W N G K N S E M T V K P F G G 120  
QY 121 N N A A L V N Q T A S D S S V M V R Q V G F G N N A T A N Q Y 151  
Db 121 G N G A A V D Q T A S N S V N V T Q V G F G N N A T A H Q Y 151

## RESULT 15

US-08-978-878-4  
; Sequence 4, Application US/08978878  
; GENERAL INFORMATION:  
; APPLICANT: NORMARK, Staffan  
; APPLICANT: OLSEN, Arne  
; TITLE OF INVENTION: FIBRONECTIN BINDING PROTEIN AS WELL AS ITS PREPARATION  
; FILE REFERENCE: 012889-081  
; CURRENT APPLICATION NUMBER: US/08/978,878  
; CURRENT FILING DATE: 1997-11-26  
; EARLIER APPLICATION NUMBER: SE 8801723-1  
; EARLIER FILING DATE: 1988-05-06

```
; EARLIER APPLICATION NUMBER: US 07/347,189
; EARLIER FILING DATE: 1989-05-04
; EARLIER APPLICATION NUMBER: US 07/789,437
; EARLIER FILING DATE: 1991-11-06
; EARLIER APPLICATION NUMBER: US 07/970,846
; EARLIER FILING DATE: 1992-11-03
; EARLIER APPLICATION NUMBER: US 08/187,865
; EARLIER FILING DATE: 1994-01-28
; EARLIER APPLICATION NUMBER: US 08/318,519
; EARLIER FILING DATE: 1994-10-05
; EARLIER APPLICATION NUMBER: US 08/495,959
; EARLIER FILING DATE: 1995-06-28
; NUMBER OF SEQ ID NOS: 10
; SOFTWARE: Patentin Ver. 2.0
; SEQ ID NO 4
; LENGTH: 151
; TYPE: PRT
; ORGANISM: Escherichia coli
US-08-978-878-4

Query Match      66.7%; Score 520; DB 13; Length 151;
Best Local Similarity 68.2%; Pred. No. 3.4e-47;
Matches 103; Conservative 21; Mismatches 27; Indels 0; Gaps 0;

Qy  1 MKLLKVAAPAAIVVSGSALAGVVPQWGGGNGHGGSSGPDYDQLVTRVVTHEMAHALQ 60
Db  1 MKLLKVAAPAAIVFSGSAGVVPQYGGGNGHGGSSGPNSELNIOYGGGNSALALQ 60

Qy  61 SDARKSETTITQGYGNGADVGGQADNSTIELTQNGFRNATIDOWNAKNSDITYGYGG 120
Db  61 TDARNSDLTITQHGCGGNGADVGGQSDSSIDLTLQRCFGNSATLDQWNGKNSMTVKQFGG 120

Qy  121 NNAAIVNOTASDSSVMVROVGFGNATANQY 151
Db  121 GNGAAVDQTASNSNVNTQVGFGNATAHQY 151
```

Search completed: August 2, 2004, 15:26:43  
Job time : 167.9 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: August 2, 2004, 14:49:38 ; Search time 17.8 Seconds  
(without alignments)  
888.146 Million cell updates/sec

Title: US-09-543-407-18  
Perfect score: 780  
Sequence: 1 MKLLKVAFAAIVVSGSALA.....DSSVMVQVGFNNATANOY 151

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 601315 seqs, 104695340 residues

Total number of hits satisfying chosen parameters: 601315

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Pending Patents AA New:

- 1: /cgn2\_6/ptodata/2/paa/PCT\_NEW\_COMB.pep.\*
- 2: /cgn2\_6/ptodata/2/paa/US06\_NEW\_COMB.pep.\*
- 3: /cgn2\_6/ptodata/2/paa/US07\_NEW\_COMB.pep.\*
- 4: /cgn2\_6/ptodata/2/paa/US08\_NEW\_COMB.pep.\*
- 5: /cgn2\_6/ptodata/2/paa/US09\_NEW\_COMB.pep.\*
- 6: /cgn2\_6/ptodata/2/paa/US10\_NEW\_COMB.pep.\*
- 7: /cgn2\_6/ptodata/2/paa/US10\_NEW\_COMB.pep.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	ID	Description
1	520	66.7	151	5	US-09-741-873C-4
2	444	56.9	131	5	US-09-741-873C-2
3	90	11.5	852	6	US-10-170-2058-3413
4	89	11.4	1627	6	US-10-170-2058-16659
5	89	11.4	1905	1	PCT-US04-09388-9
6	87	11.2	461	7	US-60-556-841-11622
7	86	11.0	841	7	US-60-565-632-7906
8	86	11.0	841	7	US-60-579-062-7906
9	85	10.9	295	6	US-10-425-115-312468
10	84	10.8	892	5	US-09-952-2678-5
11	84	10.8	892	6	US-10-872-768-5
12	84	10.8	892	6	US-10-872-769-5
13	84	10.8	1507	7	US-60-579-902-6263
14	83.5	10.7	415	5	US-09-490-324-280
15	83.5	10.7	873	5	US-09-952-2678-13
16	83.5	10.7	873	6	US-10-872-768-13
17	83.5	10.7	873	6	US-10-872-769-13
18	83	10.6	358	5	US-09-248-796A-22578
19	83	10.6	556	6	US-10-425-115-337674
20	83	10.6	573	7	US-60-565-632-7907
21	83	10.6	573	7	US-60-579-062-7907
22	82.5	10.6	434	6	US-10-045-674A-594
23	81.5	10.4	177	6	US-10-737-290-171
24	81.5	10.4	400	6	US-10-490-953-13
25	81.5	10.4	400	6	US-10-490-953-14
26	81.5	10.4	400	6	US-10-490-953-20

ALIGNMENTS

RESULT 1

US-09-741-873C-4  
; Sequence 4, Application US/09741873C  
; GENERAL INFORMATION:  
; APPLICANT: Normark, Staffan  
; TITLE OF INVENTION: Fibronectin Binding Protein As Well As Its Preparation  
; FILE REFERENCE: 012889-084  
; CURRENT APPLICATION NUMBER: US/09/741,873C  
; CURRENT FILING DATE: 2000-12-22  
; PRIOR APPLICATION NUMBER: SE 8801723-1  
; PRIOR FILING DATE: 1998-05-06  
; PRIOR APPLICATION NUMBER: US 08/978,878  
; PRIOR FILING DATE: 1997-11-26  
; PRIOR APPLICATION NUMBER: US 07/347,189  
; PRIOR FILING DATE: 1989-05-04  
; PRIOR APPLICATION NUMBER: US 07/789,437  
; PRIOR FILING DATE: 1991-11-06  
; PRIOR APPLICATION NUMBER: US 07/970,846  
; PRIOR FILING DATE: 1992-11-03  
; PRIOR APPLICATION NUMBER: US 08/187,865  
; PRIOR FILING DATE: 1994-01-28  
; PRIOR APPLICATION NUMBER: US 08/318,519  
; NUMBER OF SEQ ID NOS: 11  
; SOFTWARE: PatentIn version 3.0  
; SEQ ID NO 4  
; LENGTH: 151  
; TYPE: PRT  
; ORGANISM: Escherichia coli  
US-09-741-873C-4

27	81.5	10.4	400	6	US-10-723-981-14	Sequence 14, Appl
28	81.5	10.4	400	6	US-10-723-981-15	Sequence 15, Appl
29	81.5	10.4	424	6	US-10-045-674A-591	Sequence 591, Appl
30	81.5	10.4	533	6	US-10-045-674A-527	Sequence 527, Appl
31	81.5	10.4	599	6	US-10-491-001-22	Sequence 22, Appl
32	81	10.4	201	6	US-10-425-115-309662	Sequence 309662, Appl
33	81	10.4	412	7	US-60-565-632-7905	Sequence 7905, Ap
34	81	10.4	412	7	US-60-579-062-7905	Sequence 7905, Ap
35	80.5	10.3	258	6	US-10-425-115-303390	Sequence 303390, Ap
36	80.5	10.3	586	1	PCT-US03-24982A-317	Sequence 317, App
37	80	10.3	132	6	US-10-425-115-351875	Sequence 351875, Ap
38	80	10.3	163	5	US-09-490-324-282	Sequence 282, App
39	80	10.3	197	6	US-10-425-115-304391	Sequence 304391, Ap
40	80	10.3	374	1	PCT-US04-11210-37	Sequence 37, Appl
41	80	10.3	2319	1	PCT-US04-12717-26	Sequence 26, Appl
42	80	10.3	2319	6	US-10-831-070-26	Sequence 2, Appl
43	79.5	10.2	376	6	US-10-491-733-2	Sequence 63, Appl
44	79.5	10.2	956	6	US-10-093-037A-63	Sequence 36, Appl
45	79	10.1	284	1	PCT-US04-11210-36	

Query Match 56.7%; Score 520; DB 5; Length 151;  
Best Local Similarity 68.2%; Pred. No. 2.6e-39;  
Matches 103; Conservative 21; Mismatches 27; Indels 0; Gaps 0;

Qy	1	MKLLKVAFAAIVVSGSALA	GVVQVQVGGGNGGNGGSGGPDYDQVTRVWTHMAHALQ	60
Db	1	MKLLKVAFAAIVFSSGSA	GVVQVQVGGGNGGNGGNSPSELNIYQVGGNSALALQ	60
Qy	61	SDARKSETTTTQSGYNG	ADVGQADNISTELTQNGFRNNATIDQNAKNSDITVQYGG	120
Db	61	TDARNSDLTTQGGGNG	ADVGQGGDDSIDITQRFNGSATLDQNGKNSMTVKQFGG	120
Qy	121	NNAALVNQATSDSSVM	RVQVGFNNATANOY	151
Db	121	GNGRAVDQTASNSNV	TVQVGFNNATAHOY	151





APPLICANT: Lu, Maolong  
APPLICANT: Munyikwa, Tichifa R. I.  
APPLICANT: Roberts, James K.  
APPLICANT: Wu, Wei  
APPLICANT: Zhang, Bei  
TITLE OF INVENTION: Methods for Genetic Control of Insect Infestations in Plants and  
FILE REFERENCE: 38-21(53403)B  
CURRENT APPLICATION NUMBER: US/60/565,632  
CURRENT FILING DATE: 2004-04-27  
NUMBER OF SEQ ID NOS: 15449  
SOFTWARE: Patent in version 3.2  
SEQ ID NO 7906  
LENGTH: 841  
TYPE: PRT  
ORGANISM: Diabrotica virgifera  
FEATURE:  
NAME/KEY: misc\_feature  
LOCATION: (810)..(810)  
OTHER INFORMATION: Xaa can be any naturally occurring amino acid  
US-60-565-632-7906

Query Match 11.0%; Score 86; DB 7; Length 841;  
Best Local Similarity 24.7%; Pred. No. 17; Mismatches 47; Indels 58; Gaps 7;  
Matches 40; Conservative 17

QY 30 GNHNG--GNSSGPDYDQLVTRVVTHEMAHALQSDARK---SETTITQSGYNGA---DV 81  
Db 360 GNENGTAAGNANPD-----VQNDAAQVNDNGTAAENNGNADAAQSDND 403  
QY 82 GQAGDNSTIELTQNG-----FNATIDOWN-----AKNSDITVGYGG 120  
Db 404 GAAAEENTNADAQNGAAQGTANEANAENANADAQNDAAQNAENGAASGNADAAGT 463  
QY 121 NNAALVNQT-----ASDSSVMVRQVGFNNATAN 149  
Db 464 DNGAAENTGNADPAQNDNGCAAENSGNENGAIAENANAD 505

RESULT 8  
US-60-579-062-7906  
Sequence 7906, Application US/60579062  
GENERAL INFORMATION:  
APPLICANT: Baum, James A.  
APPLICANT: Kovalic, David K.  
APPLICANT: Larosa, Thomas J.  
APPLICANT: Lu, Maolong  
APPLICANT: Munyikwa, Tichifa R. I.  
APPLICANT: Roberts, James K.  
APPLICANT: Wu, Wei  
APPLICANT: Zhang, Bei  
TITLE OF INVENTION: Methods for Genetic Control of Insect Infestations in Plants and  
FILE REFERENCE: 38-21 (53403) C  
CURRENT APPLICATION NUMBER: US/60/579,062  
CURRENT FILING DATE: 2004-06-11  
NUMBER OF SEQ ID NOS: 41445  
SOFTWARE: Patent in version 3.2  
SEQ ID NO 7906  
LENGTH: 841  
TYPE: PRT  
ORGANISM: Diabrotica virgifera  
FEATURE:  
NAME/KEY: misc\_feature  
LOCATION: (810)..(810)  
OTHER INFORMATION: Xaa can be any naturally occurring amino acid  
US-60-579-062-7906

Query Match 11.0%; Score 86; DB 7; Length 841;  
Best Local Similarity 24.7%; Pred. No. 17;  
Matches 40; Conservative 17; Mismatches 47; Indels 58; Gaps 7;  
QY 30 GNHNG--GNSSGPDYDQLVTRVVTHEMAHALQSDARK---SETTITQSGYNGA---DV 81

PRIOR APPLICATION NUMBER: US 60/476,135  
PRIOR FILING DATE: 2003-06-04  
PRIOR APPLICATION NUMBER: US 60/476,583  
PRIOR FILING DATE: 2003-06-06  
NUMBER OF SEQ ID NOS: 40  
SOFTWARE: PERL Program  
SEQ ID NO 9  
LENGTH: 1905  
TYPE: PRT  
ORGANISM: Homo sapiens  
FEATURE:  
NAME/KEY: misc\_feature  
OTHER INFORMATION: Incyte ID No: 7525307CD1  
PCT-US04-09388-9

Query Match 11.4%; Score 89; DB 1; Length 1905;  
Best Local Similarity 24.4%; Pred. No. 23;  
Matches 39; Conservative 18; Mismatches 47; Indels 56; Gaps 7;  
QY 10 AAIVVSGSALAGV-----PWG-----GGN-----H 32  
Db 759 AAAAKSGHAWGAANQEDKSPTPPKPKSOHWGDSQNSPAMWAGGWDSSVLGH 818  
QY 33 NCGNSGPDYDQLVTRVVTHEMAHALQSDARKSETTIT-OSGYNGADVGGADNSTIE 91  
Db 819 LDGKKNGSCWD-----ADSNRSGSGWNTTTRSGSGWGNSTNTRKANPCTNWGE 867  
QY 92 LFQNGFRNATIDQNAKNSDITVGYGNNALVNQTAS 131  
Db 868 TLKPGPQQN-----WASKPQDNNVNWGG--AASVKQTGT 900

RESULT 6  
US-60-556-841-11622  
Sequence 11622, Application US/60556841  
GENERAL INFORMATION:  
APPLICANT: Abad, Mark S.  
TITLE OF INVENTION: Genes and Uses for Plant Improvement  
FILE REFERENCE: 38-21(53450)  
CURRENT APPLICATION NUMBER: US/60/556,841  
CURRENT FILING DATE: 2004-03-25  
NUMBER OF SEQ ID NOS: 12463  
SEQ ID NO 11622  
LENGTH: 461  
TYPE: PRT  
ORGANISM: Cytophaga hutchinsonii  
US-60-556-841-11622

Query Match 11.2%; Score 87; DB 7; Length 461;  
Best Local Similarity 24.2%; Pred. No. 6.7;  
Matches 40; Conservative 22; Mismatches 69; Indels 34; Gaps 7;  
QY 4 LKVAAP--RAIVVSGSALAGVWPQWGGGNGHNGG---GNSGPDYDQLVTRVVTHEMAHA 58  
Db 177 VSGAFHTWAKTDGSL-----WAGSNRNLRLGIGITTTKNAPTKVTATNWKSVYA 229  
QY 59 LQSD--ARKSETTITQSGYNGADVGGADNSTIELTQNGFRNATIDOWN----- 107  
Db 230 AGNTFAIKTDGSLWAGYNYAGVLGDTTDLRLSPVQIGTDNN-----WKTSSDNTS 284  
QY 108 AKNSDITVGYGNNALVNQTASDSSVMVRQVGFNNATANQY 151  
Db 285 FALKSDGLTWGDNNSGFGNGTTTNTLPIQIG-----TDNQW 324

RESULT 7  
US-60-565-632-7906  
Sequence 7906, Application US/60565632  
GENERAL INFORMATION:  
APPLICANT: Monsanto Technology, LLC  
APPLICANT: Baum, James A.  
APPLICANT: Kovalic, David K.  
APPLICANT: Larosa, Thomas J.  
US-60-565-632-7906

Query Match 11.0%; Score 86; DB 7; Length 841;  
Best Local Similarity 24.7%; Pred. No. 17;  
Matches 40; Conservative 17; Mismatches 47; Indels 58; Gaps 7;  
QY 30 GNHNG--GNSSGPDYDQLVTRVVTHEMAHALQSDARK---SETTITQSGYNGA---DV 81

```

Db      360  GNGEAGTAAGNNANPD-----VQNDAAQVNDNGTAAENNGNADAAQSDND 403
Qy      82    GCGADNSTIELTQNG-----FRNATIDOWN-----AKNSDITVGYGG 120
Db      404  GAAAEENTVADAQNGAAQGTANEAEENNADAAQNDAAQANENGAAENSNGNADAAQGT 463
Qy      121  NNAALVNQT-----ASDSSVMVRQVGFNNATAN 149
Db      464  DNGAAAEENTGNADPAQGNNDNGAAAEENSGNENGTAAEENANAD 505

RESULT 9
US-10-425-115-312468
; Sequence 312468, Application US/10425115
; GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J.
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
; TITLE OF INVENTION: Plants
; FILE REFERENCE: 38-21(53222)B
; CURRENT APPLICATION NUMBER: US/10/425,115
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 369326
; SEQ ID NO 312468
; LENGTH: 295
; TYPE: PRT
; ORGANISM: Zea mays
; FEATURE:
; OTHER INFORMATION: Clone ID: MFT4577_48027C.1.pep
US-10-425-115-312468

Query Match      10.9%; Score 85; DB 6; Length 295;
Best Local Similarity 27.1%; Pred. No. 6;
Matches 38; Conservative 14; Mismatches 44; Indels 44; Gaps 5;

Qy      26    MCGGNGNHGGNGSGPDYDQLVTRVVTHEMAHALQSDARKSETTITQSGVGNQADVGQGA 85
Db      138  YGGGYSGGGYSSG-GYAANGYGVGSGSGNYNASGGYSGS-----DGYNGAASGGYA 192
Qy      86    DNSTIELTQNGFRN-----NATIDQWNAKN-----SDITVGYQGNNAAL 125
Db      193  NNLSSGYNSGRYNTIGSSDGTGNGSYNPNYAGNYNTGGSSSGGTGEGFGG----- 248
Qy      126  VNQTASDSSVMVRQVGFNN 145
Db      249  -----GFGGN 253

RESULT 10
US-09-952-267B-5
; Sequence 5, Application US/09952267B
; GENERAL INFORMATION:
; APPLICANT: HANSEN, ERIC J.
; APPLICANT: AEBI, CHRISTOPH
; APPLICANT: COPE, LESLIE D.
; APPLICANT: MACIVER, ISOBEL
; APPLICANT: FISKE, MICHAEL J.
; APPLICANT: FREDEBURG, ROSS A.
; TITLE OF INVENTION: USP41 AND USP42 ANTIGENS OF MORAXELLA CATARRHALIS
; FILE REFERENCE: AMCY:024
; CURRENT APPLICATION NUMBER: US/09/952,267B
; CURRENT FILING DATE: 2001-09-12
; PRIOR APPLICATION NUMBER: US/09/336,447
; PRIOR FILING DATE: 1999-06-21
; NUMBER OF SEQ ID NOS: 98
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 5
; LENGTH: 892
; TYPE: PRT
; ORGANISM: Moraxella catarrhalis

```

```

; plueckthun, Andreas
; TITLE OF INVENTION: Protein/(Poly)peptide libraries
; NUMBER OF SEQUENCES: 373
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: James F. Haley, Jr., Esq. c/o Fish & Neave
; STREET: 1251 Avenue of the Americas
; CITY: New York
; STATE: New York
; COUNTRY: USA
; ZIP: 10021
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30 (EPO)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/490,324
; FILING DATE: 24-Jan-2000
; PRIORITY INFORMATION:
; APPLICATION NUMBER: US/09/025,769
; FILING DATE: 18-FEB-1998
; APPLICATION NUMBER: EP 95 11 3021.0
; FILING DATE: 18-AUG-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: James F. Haley, Jr., Esq.
; REGISTRATION NUMBER: 27,794
; REFERENCE/DOCKET NUMBER: MORPHO/5
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212)596-9000
; TELEFAX: (212)596-9090
; INFORMATION FOR SEQ ID NO: 280:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 415 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; SEQUENCE DESCRIPTION: SEQ ID NO: 280:
US-09-490-324-280
Query Match 10.7%; Score 83.5; DB 5; Length 415;
Best Local Similarity 29.2%; Pred. No. 12;
Matches 38; Conservative 20; Mismatches 49; Indels 23; Gaps 8
QY 27 GGG--GNHNGGSSGP-DYDQLVT----RVVTHEVAHALQSDAR-KSETTITQSGYNG 78
Db 250 GGSSEGGSGSGSGSDFYEKANANKGAMTENALQSDAKGLDSVATDYGAID 309
QY 79 ADVCGADNSFTIELQTGFENNATIDQWNAKNSDITVGQVGGNNAALVNQ-----TASDS 133
Db 310 GFIB-----VSLANG--NGATGFAGSNSQAQVGE--GDNSPLMNNFRQYLPSIQP 359
QY 134 SYMVQRQVGF 143
Db 360 SVECRPFVFVG 369
RESULT 15
US-09-952-267B-13
; Sequence 13, Application US/09952267B
; GENERAL INFORMATION:
; APPLICANT: HANSEN, ERIC J.
; APPLICANT: AEBI, CHRISTOPH
; APPLICANT: COPE, LESLIE D.
; APPLICANT: MACIVER, ISOBEL
; APPLICANT: FISKE, MICHAEL J.
; APPLICANT: FREDEBURG, ROSS A.
; TITLE OF INVENTION: USPAl AND USP2A ANTIGENS OF MORAXELLA CATARRHALIS
; FILE REFERENCE: AMCY.024
; CURRENT APPLICATION NUMBER: US/09/952,267B
; CURRENT FILING DATE: 2001-09-12
; PRIOR APPLICATION NUMBER: US/09/336,447
; PRIOR FILING DATE: 1999-06-21
; NUMBER OF SEQ ID NOS: 98

```

```
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 13
; LENGTH: 873
; TYPE: PRT
; ORGANISM: Moraxella catarrhalis
US-09-952-267B-13

Query Match      10.7%; Score 83.5; DB 5; Length 873;
Best Local Similarity 24.8%; Pred. No. 29;
Matches 40; Conservative 20; Mismatches 44; Indels 57; Gaps 10;

QY      4 LKVAFAAIVVSGSALAGVVPQWGGGNHN--GGGNS-----SGPDYDQLVTRVY 51
      | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db      34 LLIVGILGMATTASAOQTIAEQ--GKGMHSIIIGGNDNEANGDYSTVSGDYNE----- 85
      | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
QY      52 THEVAHALQSDARKSETTITQSGYNGADVGQGADNSTIELTQNGFRNATIDQWNAKNS 111
      | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db      86 -----AKGDSSTIGGGYNEAN-----GDSSTI---GGGFYN-----EAKGE 119
      | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
QY      112 DITVGOYGGNNAALVNQTASDSVMVRQVFG--NNATANQY 151
      | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db      120 SSTIG--GGDN-----NSATGMYSTIGGGDNNNSATGRY 150
```

Search completed: August 2, 2004, 15:29:52  
Job time : 17.8 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run On: August 2, 2004, 14:39:53 ; Search time 9.4 Seconds  
(without alignments)  
1545.204 Million cell updates/sec

Title: US-09-543-407-18  
Perfect score: 780  
Sequence: 1 MKLLKVAFAAIVVSGSALA.....DSSVMVQVGFNNATANQY 151

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 283366 seqs, 96191526 residues

Total number of hits satisfying chosen parameters: 283366

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : PIR 78:.\*  
1: PIR1:.\*  
2: PIR2:.\*  
3: PIR3:.\*  
4: PIR4:.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	ID	Description
1	691	88.6	151	2 JC6039	fimbrin protein ag
2	691	88.6	151	2 A10635	major curlin chain
3	523	67.1	151	2 S70788	curlin protein csg
4	501.5	64.3	152	2 D90806	curlin major subun
5	501.5	64.3	152	2 H85665	hypothetical prote
6	114	14.6	1748	2 S42136	cnjB protein - Tet
7	107.5	13.8	151	2 S70787	curlin nucleator p
8	107.5	13.8	151	2 C90806	minor curlin subun
9	107.5	13.8	151	2 G85865	curlin minor chain
10	106.5	13.7	151	2 JC6040	fimbrin protein ag
11	106.5	13.7	151	2 AH0635	nucleation compone
12	95.5	12.2	145	2 AD3143	conserved hypotet
13	95.5	12.2	145	2 H98144	hypothetical prote
14	93.5	12.0	582	2 F70675	probable PPE prote
15	93	11.9	141	2 B83143	hypothetical prote
16	93	11.9	141	2 B98145	hypothetical prote
17	92	11.8	401	2 C88571	protein C05B5.3 li
18	91.5	11.7	590	1 A45621	leishmanolysin (EC
19	91.5	11.7	2174	2 E70946	hypothetical glyci
20	91	11.7	590	2 E70946	probable PPE prote
21	90	11.5	602	1 P10221	leishmanolysin (EC
22	90	11.5	896	2 A41273	glutamate receptor
23	90	11.5	906	2 A40222	glutamate receptor
24	90	11.5	906	2 S25852	glutamate receptor
25	89.5	11.5	599	2 B42049	leishmanolysin (EC
26	88.5	11.3	1655	2 E97835	hypothetical prote
27	88	11.3	599	2 A44951	leishmanolysin (EC
28	88	11.3	906	2 S38723	glutamate receptor
29	87.5	11.2	552	2 D70604	probable PPE prote

30	87.5	11.2	1028	2 A56038	DNA-binding protei
31	87.5	11.2	1213	2 S16356	ovo protein - frui
32	87.5	11.2	1567	2 S11672	ice nucleation pro
33	87	11.2	407	2 T21956	hypothetical prote
34	86	11.0	409	2 T20847	hypothetical prote
35	85	11.0	495	2 B71360	hypothetical glyci
36	85.5	11.0	588	2 F70971	C05B5.3 protein (c
37	85	10.9	364	2 S43574	probable PPE prote
38	85	10.9	3716	2 F70825	probable PPE prote
39	85	10.9	3716	2 E70969	probable RIX fami
40	85	10.9	5188	2 B85547	hypothetical prote
41	85	10.9	5291	2 F90696	hypothetical prote
42	84.5	10.8	639	2 C42049	leishmanolysin (EC
43	84	10.8	596	2 T26667	hypothetical prote
44	84	10.8	633	2 A25473	chorion E2 protein
45	84	10.8	646	1 S19916	leishmanolysin (EC

ALIGNMENTS

RESULT 1

JC6039  
fimbrin protein agfa precursor - Salmonella enteritidis  
C:Species: Salmonella enteritidis  
C:Date: 31-Dec-1996 #sequence revision 31-Dec-1996 #text\_change 08-Oct-1999  
C:Accession: JC6039; PC6015; A44898  
R:Collinson, S.K.; Clouthier, S.C.; Doran, J.L.; Banser, P.A.; Kay, W.W.  
J. Bacteriol. 178, 662-667, 1996  
A:Title: Salmonella enteritidis agfBAC operon encoding thin, aggregative fimbrae.  
A:Reference number: JC6039; MUID:96146512; PMID:8550497  
A:Accession: JC6039  
A:Molecule type: DNA  
A:Residues: 1-151 <COL>  
A:Cross-references: GB:U03280; NID:G1184712; PIDN:AAC43599.1; PID:G1184714  
A:Accession: PC6015  
A:Molecule type: protein  
A:Residues: 21-52 <CO2>  
A:Experimental source: strain 27655-3b  
A:Note: the authors translated the codon ACG for residue 44 as Ile  
R:Collinson, S.K.; Emody, L.; Muller, K.H.; Trust, T.J.; Kay, W.W.  
J. Bacteriol. 173, 4773-4781, 1991  
A:Title: Purification and characterization of thin, aggregative fimbrae from Salmonella  
A:Reference number: A44898; MUID:91310586; PMID:1677357  
A:Contents: 27655  
A:Accession: A44898  
A>Status: Preliminary  
A:Molecule type: protein  
A:Residues: 21-33 <CO3>  
A:Note: sequence extracted from NCBI backbone (NCBIP:45936)  
C:Genetics:  
A:Gene: agfA  
C:Function:  
A:Description: major component of thin aggregative fimbrae  
A:Note: fimbrae bind to fibronectin, plasminogen, tissue plasminogen activator  
C:Keywords: fimbria  
F:1-20/Domain: signal sequence #status predicted <SIG>  
F:21-151/Product: fimbrin protein agfa #status experimental <MAT>

Query Match	88.6%	Score	691;	DB	2;	Length	151;
Best Local Similarity	90.7%	Pred. No.	9.7e-53;				
Matches	137;	Conservative	3;	Mismatches	11;	Indels	0;
Gaps	0;						
QY	1	MKLLKVAFAAIVVSGSALA	GVVPQWGGGHHNGGNSGDPDYDOLVTRVVTHEMAHALQ	60			
Db	1	MKLLKVAFAAIVVSGSALA	GVVPQWGGGHHNGGNSGDPSTLUSIYQYGSAHALQ	60			
QY	61	SDARKSETTITQSGYNGADVGQGNADNNTIELTQNGFNNATIDQWNAKNSDITVGYGG	120				
Db	61	SDARKSETTITQSGYNGADVGQGNADNNTIELTQNGFNNATIDQWNAKNSDITVGYGG	120				
QY	121	NNAALVNQTASDSSVMVQVGFNNATANQY	151				



Db 61 QADARNSLDITTHGGGNGADVGGSDSSIDLTHQFGNSATLDDWNGKDSHTVTKQFG 120

QY 120 GNNAAALVNTQASDSSVWVRQVGFNNATANQY 151

Db 121 GGNGAAVDQTASNSTVNTVQVGFNNATANQY 152

RESULT 5

H85665

hypothetical protein csgA [imported] - Escherichia coli (strain O157:H7, substrain EDL93)

C:Species: Escherichia coli

C>Date: 16-Feb-2001 #sequence\_revision 16-Feb-2001 #text\_change 14-Sep-2001

C:Accession: H85665

R:Perna, N.T.; Plunkett III, G.; Burland, V.; Mau, B.; Glaesner, J.D.; Rose, D.J.; Mayhew, L.; Grobeck, E.J.; Davis, N.W.; Lim, A.; Dimalanta, E.; Potamousis, K.; Apodaca, Nature 409, 529-533, 2001

A:Title: Genome sequence of enterohemorrhagic Escherichia coli O157:H7.

A:Reference number: A85480; MUID:21074935; PMID:11206551

A:Accession: H85665

A>Status: preliminary

A:Molecule type: DNA

A:Residues: 1-152 <STO>

A:Cross-references: GB:AE005174; NID:g12514574; PIDN:AG055788.1; GSPDB:GN00145; UWGP-Z16

A:Experimental source: strain O157:H7, substrain EDL933

C:Genetics:

A:Gene: csgA

Query Match 64.3%; Score 501.5; DB 2; Length 152;

Best Local Similarity 67.1%; Pred. No. 2.3e-36; Indels 1; Gaps 1;

Matches 102; Conservative 21; Mismatches 28;

QY 1 MKLLKVAAPAAIVSGSALAGVVPQW-GGGGNGHNGGNSGPDYDQLVTRVTVTHEMAHAL 59

Db 1 MKLLKVAAPAAIVSFGSALAGVVPQVGGGGNGHNGGNSGPNSELINIVQGGNSALAL 60

QY 60 QSDARKSETTITQSGYNGADVCGGADNSTIELTQGFNNATIDQWNAKNSDITVGVQY 119

Db 61 QADARNSLDITTHGGGNGADVGGSDSSIDLTHQFGNSATLDDWNGKDSHTVTKQFG 120

QY 120 GNNAAALVNTQASDSSVWVRQVGFNNATANQY 151

Db 121 GGNGAAVDQTASNSTVNTVQVGFNNATANQY 152

RESULT 6

S42136

cnjB protein - Tetrahymena thermophila

C:Species: Tetrahymena thermophila

C>Date: 19-Mar-1997 #sequence\_revision 18-Jul-1997 #text\_change 07-Dec-1999

C:Accession: S42136; S42135; S03650

R:Taylor, F.M.; Martindale, D.W.

submitted to the EMBL Data Library, October 1992

A:Reference number: S42136

A:Accession: S42136

A:Molecule type: DNA

A:Residues: 1-1748 <TAY>

A:Cross-references: EMBL:L03710; NID:g161751; PID:g161752

R:Taylor, F.M.; Martindale, D.W.

Nucleic Acids Res. 21, 4610-4614, 1993

A:Title: Retroviral-type zinc fingers and glycine-rich repeats in a protein encoded by c

A:Reference number: S42135; MUID:94051569; PMID:8233798

A:Accession: S42135

A:Molecule type: DNA

A:Residues: 1164-1174; 1179-1198; 1233-1252; 1285-1293; 1309; 1316-1326; 1331-1341; 1343-1

A:Cross-references: EMBL:L03710

R:Taylor, F.M.; Martindale, D.W.; Taylor, F.M.

Nucleic Acids Res. 16, 2189-2201, 1988

A:Title: Multiple introns in a conjugation-specific gene from Tetrahymena thermophila.

A:Reference number: S03650; MUID:86189811; PMID:3357771

A:Accession: S03650

A:Molecule type: DNA

A:Residues: 236-250, 'I', 252-255, 'N', 257-273 <MAR>

A:Cross-references: EMBL:X06462

C:Genetics:

A:Gene: cnjB

A:Genetic code: SGC5

A:Introns: 85/3; 136/1; 157/3; 201/2; 290/2; 327/3; 499/1; 573/2; 607/3; 708/3; 777/3; 88

C:Keywords: zinc finger

F:1164-1450/Region: glycine-rich

F:1451-1464/Region: zinc finger CCHC motif

F:1478-1491/Region: zinc finger CCHC motif

F:1501-1514/Region: zinc finger CCHC motif

F:1530-1543/Region: zinc finger CCHC motif

F:1555-1568/Region: zinc finger CCHC motif

F:1579-1592/Region: zinc finger CCHC motif

F:1602-1615/Region: zinc finger CCHC motif

F:1626-1748/Region: glycine-rich

Query Match 14.6%; Score 114; DB 2; Length 1748;

Best Local Similarity 30.5%; Pred. No. 0.11;

Matches 39; Conservative 22; Mismatches 27; Indels 40; Gaps 8;

QY 25 QWGGGNGHNGG---GNSSGPDYDQLVTRVTVTHEMAHALQSDARKSETTITQSGYNGADV 81

Db 1640 QFGGGGNSGQSGWTSSGSDMN-----CQSNVQES-TTTSGGGWS----- 1680

QY 82 GQGADNSTIELTQGFNNATIDQWNAKNSDITVGVQYNNAAALVNTQASDSSVWVRQV 141

Db 1681 -SGSGNQ-----TGGGWSN---DNQQQNEVTGGGGWSNS---NOTNNSS----- 1722

QY 142 FGNNATAN 149

Db 1723 WGSNNQAS 1730

RESULT 7

S70787

curlin nucleator protein csgB precursor - Escherichia coli (strain K-12)

N:Alternate names: csgB protein; curlin nucleation component; minor curlin protein

C:Species: Escherichia coli

C>Date: 12-Feb-1998 #sequence\_revision 20-Feb-1998 #text\_change 01-Mar-2002

C:Accession: S70787; F64846

R:Hammar, M.; Arngqvist, A.; Bian, Z.; Olsen, A.; Normark, S.

Mol. Microbiol. 18, 661-670, 1995

A:Title: Expression of two csg operons is required for production of fibronectin- and Cor

A:Reference number: S70783; MUID:96414468; PMID:8817489

A:Accession: S70787

A>Status: nucleic acid sequence not shown; translation not shown

A:Molecule type: DNA

A:Residues: 1-151 <HAM>

A:Cross-references: EMBL:X90754; NID:g1147558; PIDN:CAA62281.1; PID:g1147563

A:Experimental source: strain K12, substrain W3110

A>Note: the nucleotide sequence was submitted to the EMBL Data Library, August 1995

R:Blattner, F.R.; Plunkett III, G.; Bloch, C.A.; Perna, N.T.; Burland, V.; Riley, M.; Col

A.: Rose, D.J.; Mau, B.; Shao, Y.

Science 277, 1453-1462, 1997

A:Title: The complete genome sequence of Escherichia coli K-12.

A:Reference number: A64720; MUID:97426617; PMID:9278503

A:Accession: F64846

A>Status: nucleic acid sequence not shown; translation not shown

A:Molecule type: DNA

A:Residues: 1-151 <BLAT>

A:Cross-references: GB:AE000205; GB:U00096; NID:g1787265; PIDN:AAC74125.1; PID:g1787278;

A:Experimental source: strain K-12, substrain MG1655

C:Genetics:

A:Gene: csgB

A:Map position: 23.15

C:Function:

A:Description: minor component of wild-type curli; interaction between CsgA and CsgB trig

A>Note: curli are thin, coiled fibers expressed on the surface of Escherichia coli that i

A:kinogen; in the absence of CsgA, CsgB can self-assemble into polymers

F:1-21/Domain: signal sequence #status predicted <SIG>

F:22-151/Product: minor curlin chain #status predicted <MAR>

Query Match 13.8%; Score 107.5; DB 2; Length 151;

Best Local Similarity 35.4%; Pred. No. 0.024;  
Matches 29; Conservative 9; Mismatches 43; Indels 1; Gaps 1;

QY 70 ITQSGYNGADVQGGADNSTIELTQGFNNATIDQWNAKNSDITVQYGGNNAALVNQT 129  
DB 49 IQGAGTNSAQLRQGGSKLLAVVAQEGSSNRKIDQTDGYNL-AYIDQAGSANDASISQG 107

QY 130 ASDSSVMVRQVGFNNATANQY 151  
DB 108 AYGNTAMIIQKSGNKANTIQY 129

RESULT 8  
C90806  
minor curlin subunit precursor CsgB [imported] - Escherichia coli (strain O157:H7, subsp. C; Species: Escherichia coli  
C; Date: 18-Jul-2001 #sequence\_revision 18-Jul-2001 #text\_change 18-Jul-2001  
C; Accession: C90806  
R; Hayashi, T.; Makino, K.; Ohnishi, M.; Kurokawa, K.; Ishii, K.; Yokoyama, K.; Han, C.G.; Gasawara, N.; Yasunaga, T.; Kuhara, S.; Shiba, T.; Hattori, M.; Shinagawa, H.  
DNA Res. 9, 11-22, 2001  
A; Title: Complete genome sequence of enterohemorrhagic Escherichia coli O157:H7 and genomic islands  
A; Reference number: A99629; MUID:21156231; PMID:11258796  
A; Accession: C90806  
A; Status: preliminary  
A; Molecule type: DNA  
A; Residues: 1-151 <HAY>  
A; Cross-references: GB:BA000007; PIDN:BA834842.1; PID:gl3360879; GSPDB:GN00154  
A; Experimental source: strain O157:H7, substrain RMD 0509952  
C; Genetics:  
A; Gene: ECs1419

Query Match 13.8%; Score 107.5; DB 2; Length 151;  
Best Local Similarity 35.4%; Pred. No. 0.024;  
Matches 29; Conservative 9; Mismatches 43; Indels 1; Gaps 1;

QY 70 ITQSGYNGADVQGGADNSTIELTQGFNNATIDQWNAKNSDITVQYGGNNAALVNQT 129  
DB 49 IQGAGTNSAQLRQGGSKLLAVVAQEGSSNRKIDQTDGYNL-AYIDQAGSANDASISQG 107

QY 130 ASDSSVMVRQVGFNNATANQY 151  
DB 108 AYGNTAMIIQKSGNKANTIQY 129

RESULT 9  
G85665  
curlin minor chain precursor, CsgA homolog [imported] - Escherichia coli (strain O157:H7, subsp. C; Species: Escherichia coli  
C; Date: 16-Feb-2001 #sequence\_revision 16-Feb-2001 #text\_change 14-Sep-2001  
C; Accession: G85665  
R; Perna, N.T.; Plunkett III, G.; Burland, V.; Mau, B.; Glasner, J.D.; Rose, D.J.; Mayhew  
Miller, L.; Grothbeck, E.J.; Davis, N.W.; Lim, A.; Dimalanta, E.; Potamousis, K.; Apodaca,  
Nature 409, 529-533, 2001  
A; Title: Genome sequence of enterohemorrhagic Escherichia coli O157:H7.  
A; Reference number: A85480; MUID:21074935; PMID:11206551  
A; Accession: G85665  
A; Status: preliminary  
A; Molecule type: DNA  
A; Residues: 1-151 <STO>  
A; Cross-references: GB:AE005174; NID:G12514573; PIDN:AA655787.1; GSPDB:GN00145; UWGP:Z16  
A; Experimental source: strain O157:H7, substrain EDL933  
C; Genetics:  
A; Gene: csgB

Query Match 13.8%; Score 107.5; DB 2; Length 151;  
Best Local Similarity 35.4%; Pred. No. 0.024;  
Matches 29; Conservative 9; Mismatches 43; Indels 1; Gaps 1;

QY 70 ITQSGYNGADVQGGADNSTIELTQGFNNATIDQWNAKNSDITVQYGGNNAALVNQT 129  
DB 49 IQGAGTNSAQLRQGGSKLLAVVAQEGSSNRKIDQTDGYNL-AYIDQAGSANDASISQG 107

QY 130 ASDSSVMVRQVGFNNATANQY 151  
DB 108 AYGNTAMIIQKSGNKANTIQY 129

RESULT 10  
JC6040  
fimbrin protein egfB precursor - Salmonella enteritidis  
C; Species: Salmonella enteritidis  
C; Date: 31-Dec-1996 #sequence\_revision 31-Dec-1996 #text\_change 08-Oct-1999  
C; Accession: JC6040  
R; Collinson, S.K.; Clouthier, S.C.; Doran, J.L.; Baner, P.A.; Kay, W.W.  
J. Bacteriol. 178, 662-667, 1996  
A; Title: Salmonella enteritidis egfBAC operon encoding thin, aggregative fimbrinae.  
A; Reference number: JC6039; MUID:96146512; PMID:8550497  
A; Accession: JC6040  
A; Molecule type: DNA  
A; Residues: 1-151 <COL>  
A; Cross-references: GB:U43280; NID:gl184712; PIDN:AA043598.1; PID:gl1184713  
A; Experimental source: strain 276755-3b  
C; Genetics:  
A; Gene: agfB  
C; Function:  
A; Description: minor component of thin aggregative fimbrinae  
A; Note: fimbrinae bind to fibronectin, plasminogen, tissue plasminogen activator  
C; Keywords: fimbrina  
P; 1-21/Domain: signal sequence #status predicted <SIG>  
P; 22-151/Product: fimbrin protein agfB #status predicted <MAT>

Query Match 13.7%; Score 106.5; DB 2; Length 151;  
Best Local Similarity 30.7%; Pred. No. 0.029;  
Matches 35; Conservative 17; Mismatches 43; Indels 19; Gaps 5;

QY 55 MAHALQSDARKSE-----TTITQSGYNGADVQGGADNST-----TELTONGFR 98  
DB 18 IATATNYDLARSEYNFAVNELSKSSFNQAIIQGVGTDSARVRQEGSKLLSVISQEGN 77

QY 99 NNATIDQWNAKNSDIT-VQYGGNNAALVNQTASDSSVMVRQVGFNNATANQY 151  
DB 78 NRAKVDQ--AGNYNFAVIEQTGNANDASISQSAVYNSAAIIQKSGNKANTIQY 129

RESULT 11  
AH0635  
nucleation component of curlin monomers [imported] - Salmonella enterica subsp. enterica  
C; Species: Salmonella enterica subsp. enterica serovar Typhi  
A; Note: this species has also been called Salmonella typhi  
C; Date: 09-Nov-2001 #sequence\_revision 09-Nov-2001 #text\_change 18-Nov-2002  
C; Accession: AH0635  
R; Parkhill, J.; Dougan, G.; James, K.D.; Thomson, N.R.; Pickard, D.; Wain, J.; Churcher,  
th, T.; Connor, P.; Cronin, A.; Davis, P.; Davies, R.M.; Dowd, L.; White, N.; Farrar,  
S.; Moulie, S.; O'Gaora, P.  
Nature 413, 848-852, 2001  
A; Authors: Parry, C.; Quail, M.; Rutherford, K.; Simmonds, M.; Skelton, J.; Stevens, K.;  
A; Title: Complete genome sequence of a multiple drug resistant Salmonella enterica serovar  
A; Reference number: AB0502; MUID:21534947; PMID:11677608  
A; Accession: AH0635  
A; Status: preliminary  
A; Molecule type: DNA  
A; Residues: 1-151 <PAR>  
A; Cross-references: GB:AL513382; PIDN:CAD08267.1; PID:gi6502314; GSPDB:GN00176  
C; Genetics:  
A; Gene: STY1180

Query Match 13.7%; Score 106.5; DB 2; Length 151;  
Best Local Similarity 30.7%; Pred. No. 0.029;  
Matches 35; Conservative 17; Mismatches 43; Indels 19; Gaps 5;

QY 55 MAHALQSDARKSE-----TTITQSGYNGADVQGGADNST-----TELTONGFR 98  
DB 18 IATATNYDLARSEYNFAVNELSKSSFNQAIIQGVGTDSARVRQEGSKLLSVISQEGN 77

QY 99 NNATIDQWNAKNSDIT-VQYGGNNAALVNQTASDSSVMVRQVGFNNATANQY 151  
DB 78 NRAKVDQ--AGNYNFAVIEQTGNANDASISQSAVYNSAAIIQKSGNKANTIQY 129



```

Qy 50 VVTHEMAHALQSDARKSETTITQSGYGNAGDVGOCADNSTIELTQNGFRNNATIDQWNAK 109
      :|||
Db 61 IVGH-----QYGRHNL SAVGECHDNYGSTTQNGNRVAGI----- 96
      :|||
Qy 110 NSDITVQYGGNNAALVNQTASDSVMVVRQVGFNNATAFQ 150
      :|||
Db 97 -----GQFGSNHTTILTDQDGNNTAAGVQVGRGCSANVSQ 131
      :|||

RESULT 14
F70675
probable PPE protein - Mycobacterium tuberculosis (strain H37RV)
C:Species: Mycobacterium tuberculosis
C:Date: 17-Jul-1998 #sequence_revision 17-Jul-1998 #text_change 15-Sep-2003
C:Accession: F70675
R:Cole, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon, S.;
Rajandream, M.A.; Davies, R.; Devlin, K.; Fellwell, T.; Gentles, S.; Hamlin, N.; Holroyd, S.;
Connor, R.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S.
Nature 393, 537-544, 1998
A:Authors: Sgarbes, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.
A:Title: Deciphering the biology of Mycobacterium tuberculosis from the complete genome
A:Reference number: A70500; MUID:98295987; PMID:9634230
A:Accession: F70675
A:Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-582 <COL>
A:Cross-references: GB:282098; GB:AL123456; NID:93261664; PIDN:CAB05045.1; PID:e280737;
A:Experimental source: strain H37rv
C:Genetics:
A:Gene: PPE

Query Match 12.0%; Score 93.5; DB 2; Length 582;
Best Local Similarity 27.1%; Pred. No. 1.8;
Matches 38; Conservative 16; Mismatches 71; Indels 15; Gaps 6;

Qy 23 VPQWGGG--GNHNGGNSGSPDYDLVTRVWTHEMAHALQSDARKSETTITQSGYGNAGD 80
      :|||
Db 175 LPNLGIGNTGNNALGGGNTG-DLNTGNGNIGNTNLGSGNRRGDANLGSIGNIGNSNVG-GGN 232
      :|||
Qy 81 VQG---CADNSTIEL-----TQNGFRNNATIDQWNAKSDITVQYGGNNAALVNQTASD 132
      :|||
Db 233 VGNNTGSGNRRAGLPGSGVNGVNGNLGNSNLGSGNTGNSVGFNTGNNNVGTGNAAGSN 292
      :|||

```

Db 293 IGAGNTGSSNWGFGNNIGN 312

RESULT 15

AB3143 [AB3143] - Agrobacterium tumefaciens (strain C58, Dupont)  
hypothetical protein Atu4766 [imported]  
C:Species: Agrobacterium tumefaciens  
C:Date: 11-Jan-2002 #sequence\_revision 11-Jan-2002 #text\_change 18-Nov-2002  
C:Accession: AB3143  
R:Wood, D.W.; Setubal, J.C.; Kaul, R.; Monks, D.; Chen, L.; Wood, G.E.; Chen, Y.; Woo, L.  
erage, G.; Gillet, W.; Grant, C.; Guenther, D.; Kutyavin, T.; Levy, R.; Li, M.; McClellan,  
; Karp, P.; Romero, P.; Zhang, S.  
Science 294, 2317-2323, 2001  
A:Authors: Yoo, H.; Tao, Y.; Biddle, P.; Jung, M.; Krespan, W.; Perry, M.; Gordon-Kamm,  
ster, E.W.  
A:Title: The Genome of the Natural Genetic Engineer Agrobacterium tumefaciens C58.  
A:Reference number: AB2577; MUID:21608550; PMID:11743193  
A:Accession: AB3143  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-141 <KUR>  
A:Cross-references: GB:AE008689; PIDN:AAL45560.1; PID:gi17743275; GSPDB:GN00187  
C:Genetics:  
A:Gene: Atu4766  
A:Map position: linear chromosome

Query March 11 98: Score 93: DB 2: Length 141:



GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: August 2, 2004, 14:36:12 ; Search time 5.3 Seconds  
(without alignments)  
1483.508 Million cell updates/sec

Title: US-09-543-407-18

Perfect score: 780  
Sequence: 1 MLLKVAFAAIVVSGSALA.....DSSVMVRQVGFNNATANY 151

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 141681 seqs, 52070155 residues

Total number of hits satisfying chosen parameters: 141681

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : SwissProt\_42.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	631	88.6	151	1 CSGA_SALTY	P55225 salmonella
2	523	67.1	151	1 CSGA_ECOLI	P28307 escherichia
3	501.5	64.3	152	1 CSGA_ECOLI	Q93024 escherichia
4	107.5	13.8	151	1 CSGB_ECOLI	P39828 escherichia
5	106.5	13.7	151	1 CSGB_SALTY	Q827m3 salmonella
6	106.5	13.7	151	1 CSGB_SALTY	P55226 salmonella
7	92	11.8	401	1 YK03_CAEEL	P43291 caenorhabdi
8	92	11.8	156	1 OMPB_RICJA	O06553 r outer mem
9	91.5	11.7	590	1 GP63_LEIDO	P23223 leishmania
10	91	11.7	1093	1 PER_DROWI	Q03297 drosophila
11	90	11.5	602	1 GP63_LEIMA	P08148 leishmania
12	90	11.5	906	1 GLR1_HUMAN	P42261 homo sapien
13	88.5	11.3	1655	1 OMPB_RICCN	Q9kka3 r outer mem
14	88	11.3	559	1 GP63_LEICH	P15706 leishmania
15	87.5	11.2	1028	1 OVO_DROME	P15121 drosophila
16	87.5	11.2	1567	1 ICEN_XANCT	P18127 xanthomonas
17	86	11.0	485	1 Y136_TREPA	O83172 treponema p
18	85	11.0	493	1 GATA_THREMA	Q922k7 rhizobium m
19	84	10.8	646	1 GP63_LEIME	P43150 leishmania
20	84	10.8	907	1 GLR1_MOUSE	P23818 mus musculus
21	84	10.8	907	1 GLR1_RAT	P19490 rattus norv
22	83	10.6	720	1 KREB_YEAST	P32486 saccharomyc
23	83	10.6	947	1 SECA_STRCO	P55021 streptomyce
24	82	10.5	342	1 OMPC_RAHQO	O33507 rannellia eq
25	81.5	10.4	424	1 COAA_BPFD	P03661 bacterioph
26	81.5	10.4	424	1 COAA_BPM13	P03662 bacterioph
27	81.5	10.4	493	1 GATA_AGR5	Q8uf88 agrobacteri
28	81	10.4	365	1 ROAL_DROME	P07909 drosophila
29	81	10.4	576	1 DEAF_DROME	Q24180 drosophila
30	81	10.4	678	1 ICEN_MICTU	Q10778 mycobacteri
31	81	10.4	1034	1 ICEN_PANAN	Q47879 pantoea ana
32	80.5	10.3	641	1 IMD_ARTGO	Q44052 arthrobacte
33	80.5	10.3	1185	1 MAPX_DROME	P23226 drosophila

34	80	10.3	347	1 MSA2_PLAF2	Q03646 plasmodium
35	79.5	10.2	940	1 SECA_STRGR	P95759 streptomyce
36	79.5	10.2	1300	1 L20K_RICRI	P14914 rickettsia
37	79.5	10.2	1654	1 OMPB_RICRI	Q53047 r outer mem
38	79	10.1	199	1 AC22_TENMO	P28368 tensbrio mo
39	79	10.1	1140	1 YM96_YEAST	Q04893 saccharomyc
40	78.5	10.1	704	1 LIPM_NEIMA	P57037 neisseria m
41	78.5	10.1	1293	1 MLE_DROME	P24785 drosophila
42	78	10.0	485	1 YB47_MYCPN	P75139 mycoplasma
43	78	10.0	588	1 CAR7_CANAL	P43096 candida alb
44	78	10.0	1258	1 ICEN_ERWHE	P16239 erwinia her
45	78	10.0	2003	1 YDBA_ECOLI	P33666 escherichia

ALIGNMENTS

RESULT 1					
ID	CSGA_SALTY	STANDARD;	PRT;	151	AA.
AC	P55225;				
DT	01-OCT-1996 (Rel. 34, Created)				
DT	01-OCT-1996 (Rel. 34, Last sequence update)				
DT	10-OCT-2003 (Rel. 42, Last annotation update)				
DE	Major curlin subunit precursor (Fimbrin SEF17).				
GN	CSGA OR AGFA OR STM1144 OR STY1181 OR TI776.				
OS	Salmonella typhimurium,				
OS	Salmonella typhi, and				
OS	Salmonella enteritidis.				
OC	Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;				
OC	Enterobacteriaceae; Salmonella.				
OX	NCBI_TaxID=602, 601, 592;				
RN	[1]				
RP	SEQUENCE FROM N.A.				
RC	SPECIES=S typhimurium; STRAIN=SR-11;				
EX	MEDLINE=98117058; PubMed=9457880;				
RA	Rowling U., Bian Z., Hammar M., Sierralta W.D., Normark S.;				
RT	"Curli fibers are highly conserved between Salmonella typhimurium and				
RT	Escherichia coli with respect to operon structure and regulation."				
RL	J. Bacteriol. 180:722-731(1998).				
[2]					
RP	SEQUENCE FROM N.A.				
RC	SPECIES=S typhimurium; STRAIN=LT2 / SGSC1412 / ATCC 700720;				
EX	MEDLINE=21534948; PubMed=1677609;				
RA	McClelland M., Sanderson K.E., Spieth J., Clifton S.W., Latreille P.,				
RA	Courtney L., Porwollik S., Ali J., Dante M., Du F., Hou S., Layman D.,				
RA	Leonard S., Nguyen C., Scott K., Holmes A., Grewal N., Mulvaney E.,				
RA	Ryan E., Sun H., Florea L., Miller W., Stoneking T., Nhan M.,				
RA	Waterston R., Wilson R.K.;				
RT	"Complete genome sequence of Salmonella enterica serovar Typhimurium				
RT	LT2."				
RL	Nature 413:852-856(2001).				
[3]					
RP	SEQUENCE FROM N.A.				
RC	SPECIES=S typhi; STRAIN=CT18;				
EX	MEDLINE=21534947; PubMed=11677608;				
RA	Parkhill J., Dougan G., James K.D., Thomson N.R., Pickard D., Wain J.,				
RA	Churcher C., Mungall K.L., Bentley S.D., Holden M.T.G., Sebahia M.,				
RA	Baker S., Basham D., Brooks K., Chillingworth T., Conerton P.,				
RA	Cronin A., Davis P., Davies R.M., Dowd L., White N., Farrar J.,				
RA	Feltwell T., Hamlin N., Haque A., Hien T.T., Holroyd S., Jagels K.,				
RA	Krogh A., Larsen T.S., Leather S., Moule S., O'Gaora P., Parry C.,				
RA	Quail M.A., Rutherford K., Simmonds M., Skelton J., Stevens K.,				
RA	Whitehead S., Barrall B.G.;				
RT	"Complete genome sequence of a multiple drug resistant Salmonella				
RT	enterica serovar Typhi CT18."				
RL	Nature 413:848-852(2001).				
[4]					
RP	SEQUENCE FROM N.A.				
RC	SPECIES=S typhi; STRAIN=TY2 / ATCC 700931;				
EX	MEDLINE=22531367; PubMed=12644504;				
RA	Deng W., Liou S.-R., Plunkett G. III, Mayhew G.F., Rose D.J.,				
RA	Burland V., Kodoyianni V., Schwartz D.C., Blattner F.R.;				

RT "Comparative genomics of Salmonella enterica serovar Typhi strains Ty2  
 RL J. Bacteriol. 185:2330-2337(2003).  
 RN [5]  
 RP SEQUENCE FROM N.A.  
 RC SPECIES-S. enteritidis; STRAIN=27655-3B;  
 RX MEDLINE=96146512; PubMed=8550497;  
 RA Collinson S.K., Clouthier S.C., Doran J.L., Baner P.A., Kay W.W.,  
 RT "Salmonella enteritidis agfBAC operon encoding thin, aggregative  
 RL fimbriae";  
 RL J. Bacteriol. 178:662-667(1996).  
 RN [6]  
 RP SEQUENCE OF 21-151 FROM N.A.  
 RC SPECIES-S. enteritidis; STRAIN=27655-3B;  
 RX MEDLINE=94013373; PubMed=8104955;  
 RA Doran J.L., Collinson S.K., Burian J., Santos G., Todd E.C.D.,  
 RT "Purification and characterization of thin, aggregative fimbriae from  
 RL Salmonella enteritidis";  
 RL J. Bacteriol. 173:4773-4781(1991).  
 CC -1- FUNCTION: CURLIN IS THE STRUCTURAL SUBUNIT OF THE CURLI. CURLI ARE  
 CC COILED SURFACE STRUCTURES THAT ASSEMBLE PREFERENTIALLY AT GROWTH  
 CC TEMPERATURES BELOW 37 DEGREES CELSIUS. CURLI CAN BIND TO  
 CC FIBRONECTIN.  
 CC -1- SIMILARITY: BELONGS TO THE CSGA/CSGB FAMILY.  
 CC  
 CC This SWISS-PROT entry is copyrighted. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation  
 CC at the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC  
 CC EMBL; AJ002301; CAA05317.1; -  
 CC EMBL; A5008749; RAL20074.1; -  
 CC EMBL; ALG27269; CAD08268.1; -  
 CC EMBL; A5016840; AAC069399.1; -  
 CC EMBL; U43280; AAC43599.1; -  
 CC FIC; JC6039; JC6039.  
 CC StyGene; SGI0608; csGA.  
 CC FimB; Signal; Complete proteome.  
 FT SIGNAL 1 20  
 FT CHAIN 21 151 MAJOR CURLIN SUBUNIT.  
 FT CONFLICT 134 151 SVMVRQVGFNNATANYQ -> DSTVQVAS (IN  
 FT REF. 6).  
 FT SEQUENCE 151 AA; 15305 MW; B7DAC0D16B621359 CRC64;  
 Query Match 88.6%; Score 691; DB 1; Length 151;  
 Best Local Similarity 90.7%; Pred. No. 5.3e-53;  
 Matches 137; Conservative 3; Mismatches 11; Indels 0; Gaps 0;  
 QY 1 MKLLKVAAPAAIIVVSGSALAGVVPQGGNGNNGGNSGGPDYDQVTRVTHEMAHALQ 60  
 DB 1 MKLLKVAAPAAIIVVSGSALAGVVPQGGNGNNGGNSGGPDSTLSITQYGSANAALALQ 60  
 QY 61 SPARKSETTITGSGYNGGADVGGADNSITLTQNGFRNATIDOWNAKNSDITVQYGG 120  
 DB 61 SPARKSETTITGSGYNGGADVGGADNSITLTQNGFRNATIDOWNAKNSDITVQYGG 120  
 QY 121 NNAALVNTASDSSVMVRQVGFNNATANYQ 151  
 DB 121 NNAALVNTASDSSVMVRQVGFNNATANYQ 151

RESULT 2  
 CSGA\_ECOLI STANDARD; PRT; 151 AA.  
 AC P28307; 1992 (Rel. 24, Created)  
 DT 01-DEC-1992 (Rel. 34, Last sequence update)  
 DT 01-OCT-1996 (Rel. 41, Last annotation update)  
 DE Major curlin subunit precursor.  
 GN CSGA OR B1042.  
 OS Escherichia coli.  
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;  
 OC Enterobacteriaceae; Escherichia.  
 OX NCBI\_TaxID=562;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=K12 / W3110;  
 RX MEDLINE=93211294; PubMed=8459772;  
 RA Olsen A., Arqvist A.;  
 RT "The RpoS sigma factor relieves H-NS-mediated transcriptional  
 RL repression of csGA, the subunit gene of fibronectin-binding curli in  
 RL Escherichia coli";  
 RL Mol. Microbiol. 7:523-536(1993).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=K12 / MC4100;  
 RX MEDLINE=96414469; PubMed=8817489; Olsen A., Normark S.;  
 RA Hammar M., Arqvist A., Bian Z.,  
 RT "Expression of two csG operons is required for production of  
 RL fibronectin- and congo red-binding curli polymers in Escherichia coli  
 RL K-12";  
 RL Mol. Microbiol. 18:661-670(1995).  
 RN [3]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=K12 / MG1655;  
 RX MEDLINE=97426617; PubMed=9278503;  
 RA Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,  
 RA Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,  
 RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,  
 RA Mau B., Shao Y.;  
 RT "The complete genome sequence of Escherichia coli K-12";  
 RL Science 277:1453-1474(1997).  
 RN [4]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=K12;  
 RX MEDLINE=97061202; PubMed=8905232;  
 RA Oshima T., Aiba H., Baba T., Fujita K., Hayashi K., Honjo A.,  
 RA Ikemoto K., Inada T., Itoh T., Kajihara M., Kanai K., Kashimoto K.,  
 RA Kimura S., Kitagawa M., Makino K., Masuda S., Miki T., Mizobuchi K.,  
 RA Mori H., Motomura K., Nakamura Y., Nishimoto H., Nishio Y., Saito N.,  
 RA Sampei G., Seki Y., Tagami H., Takemoto K., Wada C., Yamamoto Y.,  
 RA Yano M., Horiuchi T.;  
 RT "A 718-kb DNA sequence of the Escherichia coli K-12 genome  
 RL corresponding to the 12.7-28.0 min region on the linkage map";  
 RL DNA Res. 3:137-155(1996).  
 RN [5]  
 RP SEQUENCE OF 21-40.  
 RC STRAIN=K12 / YMEI;  
 RX MEDLINE=93023873; PubMed=1357528;  
 RA Arqvist A., Olsen A., Pfeifer J., Russell D.G., Normark S.;  
 RT "The Crl protein activates cryptic genes for curli formation and  
 RL fibronectin binding in Escherichia coli HB101";  
 RL Mol. Microbiol. 6:2443-2452(1992).  
 RN [6]  
 RP SEQUENCE OF 21-31.  
 RX MEDLINE=91310586; PubMed=1677357;  
 RA Collinson S.K., Emeody L., Trust T.J., Kay W.W.;  
 RT "Purification and characterization of thin, aggregative fimbriae from  
 RL Salmonella enteritidis";  
 RL J. Bacteriol. 173:4773-4781(1991).  
 CC -1- FUNCTION: CURLIN IS THE STRUCTURAL SUBUNIT OF THE CURLI. CURLI ARE  
 CC COILED SURFACE STRUCTURES THAT ASSEMBLE PREFERENTIALLY AT GROWTH  
 CC TEMPERATURES BELOW 37 DEGREES CELSIUS. CURLI CAN BIND TO  
 CC FIBRONECTIN.

```

[3]
RN  SEQUENCE FROM N.A.
RC  STRAIN=O157:H7 / RIMD 0509952;
RX  MEDLINE=2115231; PubMed11258796;
RA  Hayashi T., Makino K., Onishi M., Kurokawa K., Ishii K., Yokoyama K.,
RA  Han C.-G., Ohtsubo E., Nakayama K., Murata T., Tanaka M., Tobe T.,
RA  Iida T., Takami H., Honda T., Sasakawa C., Ogasawara N., Yasunaga T.,
RA  Kuhara S., Shiba T., Hattori M., Shinagawa H.;
RT  "Complete genome sequence of enterohemorrhagic Escherichia coli
RT  O157:H7 and genomic comparison with a laboratory strain K-12.";
RL  DNA Res. 8:11-22(2001).
CC  -!- FUNCTION: CURLIN IS THE STRUCTURAL SUBUNIT OF THE CURLI. CURLI ARE
CC  COILED SURFACE STRUCTURES THAT ASSEMBLE PREFERENTIALLY AT GROWTH
CC  TEMPERATURES BELOW 37 DEGREES CELSIUS. CURLI CAN BIND TO
CC  FIBRONECTIN.
CC  -!- SIMILARITY: BELONGS TO THE CSGA/CSGB FAMILY.
CC  THIS SWISS-PROT entry is copyright. It is produced through a collaboration
CC  between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC  the European Bioinformatics Institute. There are no restrictions on its
CC  use by non-profit institutions as long as its content is in no way
CC  modified and this statement is not removed. Usage by and for commercial
CC  entities requires a license agreement (See http://www.isb-sib.ch/announcement
CC  or send an email to license@isb-sib.ch).
CC  -----
CC  EMBL; AF275733; AAK53212.1; -.
CC  DR  EMBL; AF005315; AAG55788.1; -.
CC  DR  EMBL; AF002554; BAB34843.1; -.
CC  DR  PIR; D90806; D90806.
CC  DR  PIR; H85605; H85605.
CC  DR  Fimbrin; Signal; Complete proteome.
KW  SIGNAL
FT  CHAIN 1 20 BY SIMILARITY
FT  CHAIN 21 152 MAJOR CURLIN SUBUNIT.
SQ  SEQUENCE 152 AA; 15099 MW; BE2D2D94DD591243 CRC64;
Query Match 64.3%; Score 501.5; DB 1; Length 152;
Best Local Similarity 67.1%; Pred. No. 1.2e-36;
Matches 102; Conservative 21; Mismatches 28; Indels 1; Gaps 1
Qy 1 MKLLKVAFAIAIVSGSALAGVFPQW-GGGNHGGGNSGPDYDQIVTRVVTHEMAHAL 59
Db 1 MKLLKVAFAIAIVSGSALAGVFPQYGGGNGHGGGNSGPNSELNYYOYGGGNSALAL 60
Qy 60 QSDARKSETTITQSYNGADVGQADNSTIELTQGFNRNATIDQWNAFNSDITVQYQY 119
Db 61 QADARNSDLTITQHGNGADVGQSDSSIDLITQRFNGNSALTDQWNGKDSHTVTKQFG 120
Qy 120 GNNAAALVNQTASDSSVMYRVQVFGNNATANYQ 151
Db 121 GGGAADVQDTASNSTVNVTVQVFGNNATAHQY 152
RESULT 4
CSGB ECOLI STANDARD; PRT; 151 AA.
AC -CSGB ECOLI
AD P39828;
DT 01-FEB-1995 (Rel. 31, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Minor curlin subunit precursor.
DE CSGB OR B1041 OR Z1675 OR ECS1419.
DE Escherichia coli, and
DE Escherichia coli O157:H7.
OS Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Escherichia.
CX NBI_TaxID=562, 83334;
[1]
RN  SEQUENCE FROM N.A.
RC  STRAIN=K12 / MC4100;
RX  MEDLINE=96414468; PubMed=8817489;
RA  Hammar M., Arnoqvist A., Bian Z., Olsen A., Normark S.;
RT  "Expression of two csg operons is required for production of
RT  fibronectin- and Congo red-binding curli polymers in Escherichia coli

```

RT K-12.";  
 RL Mol. Microbiol. 18:661-670(1995).  
 RP [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=K12 / MG1655;  
 RC MEDLINE=97426617; PubMed=9278503;  
 RX Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,  
 RA Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,  
 RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,  
 RA Mau B., Shao Y.;  
 RT "The complete genome sequence of *Escherichia coli* K-12.";  
 RL Science 277:1453-1474(1997).  
 RN [3]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=K12;  
 RC MEDLINE=97061202; PubMed=8905232;  
 RX Oshima T., Aiba H., Baba T., Fujita K., Hayashi K., Honjo A.,  
 RA Ikemoto K., Inada T., Itoh T., Kajihara M., Kanai K., Kashimoto K.,  
 RA Kimura S., Kitagawa M., Makino K., Masuda S., Miki T., Mizobuchi K.,  
 RA Mori H., Motomura K., Nakamura Y., Nishimoto H., Nishio Y., Saito N.,  
 RA Sempel G., Seki Y., Tagami H., Takemoto K., Wada C., Yamamoto Y.,  
 RA Yano M., Horiuchi T.;  
 RT "A 718-kb DNA sequence of the *Escherichia coli* K-12 genome  
 corresponding to the 12.7-28.0 min region on the linkage map.";  
 RL DNA Res. 3:137-155(1996).  
 RN [4]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=O157:H7 / EDL933 / ATCC 700927;  
 RX MEDLINE=21074935; PubMed=11206551;  
 RA Perna N.T., Plunkett G. III, Burland V., Mau B., Glasner J.D.,  
 RA Rose D.J., Mayhew G.F., Evans P.S., Gregor J., Kirkpatrick H.A.,  
 RA Posfai G., Hackett J., Klink S., Boutin A., Shao Y., Miller L.,  
 RA Grobeck E.J., Davis N.W., Lim A., Dimalanta E.T., Pocamorus K.,  
 RA Apodaca J., Anantharaman T.S., Lin J., Yen G., Schwartz D.C.,  
 RA Welch R.A., Blattner F.R.;  
 RT "Genome sequence of enterohaemorrhagic *Escherichia coli* O157:H7.";  
 RL Nature 409:529-533(2001).  
 RN [5]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=O157:H7 / RIMD 0509952;  
 RX MEDLINE=21156231; PubMed=11258796;  
 RA Hayashi T., Makino K., Onishi M., Kurokawa K., Ishii K., Yokoyama K.,  
 RA Han C.-G., Ohtsubo E., Nakayama K., Murata T., Tanaka M., Tobe T.,  
 RA Iida T., Takami H., Honda T., Sasakawa C., Ogasawara N., Yasunaga T.,  
 RA Kuhara S., Shiba T., Hattori M., Shinagawa H.;  
 RT "Complete genome sequence of enterohaemorrhagic *Escherichia coli*  
 O157:H7 and genomic comparison with a laboratory strain K-12.";  
 RL DNA Res. 8:11-22(2001).  
 RN [6]  
 RP SEQUENCE OF 1-21 FROM N.A.  
 RC STRAIN=K12;  
 RX MEDLINE=95157246; PubMed=7854117;  
 RA Arngvist A., Olsen A., Normark S.;  
 RT "Sigma S-dependent growth-phase induction of the *csgBA* promoter in  
*Escherichia coli* can be achieved in vivo by sigma 70 in the absence  
 of the nucleoid-associated protein H-NS.";  
 RL Mol. Microbiol. 13:1021-1032(1994).  
 CC -!- FUNCTION: CURLIN IS THE STRUCTURAL SUBUNIT OF THE CURLI. CURLI ARE  
 CC COILED SURFACE STRUCTURES THAT ASSEMBLE PREFERENTIALLY AT GROWTH  
 CC TEMPERATURES BELOW 37 DEGREES CELSIUS. CURLI CAN BIND TO  
 CC FIBRONECTIN. THE MINOR SUBUNIT IS THE NUCLEATION COMPONENT OF  
 CC CURLIN MONOMERS.  
 CC -!- SIMILARITY: BELONGS TO THE CSGA/CSGB FAMILY.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 CC EMBL, X90754; CAA62281.1; -.

DR EMBL; AE000205; AAC74125.1; -.  
 DR EMBL; D90741; BAA35831.1; -.  
 DR EMBL; AP005315; AAG55787.1; -.  
 DR EMBL; AP002554; BA934842.1; -.  
 DR PIR; C90806; C90806.  
 DR PIR; G85665; G85665.  
 DR PIR; S70787; S70787.  
 DR EcoGene; EG12621; csgB.  
 KW Fibria; Signal; Complete proteome.  
 FT SIGNAL 1 21 POTENTIAL.  
 FT CHAIN 22 151 MINOR CURLIN SUBUNIT.  
 SQ SEQUENCE 151 AA; 15882 MW; B18D26B964014B8 CRC64;  
 [3]  
 Query Match 13.8%; Score 107.5; DB 1; Length 151;  
 Best Local Similarity 35.4%; Pred. No. 0.012;  
 Matches 29; Conservative 9; Mismatches 43; Indels 1; Gaps 1;  
 QY 70 ITOSGYNGADVGGADNSITELTONGFRNATIDQWNAKNSDITVGOYGNNAALVNOT 129  
 DB 49 IGQAGTNSAQLOGGSKLLAVWAQGSNRAKIDQGYNL-AYIDQAGSANDASISQG 107  
 QY 130 ASDSSVMVROVGFNGNNATNQY 151  
 DB 108 AYGNTAMIIQKSGCNKANTQY 129  
 RESULT 5  
 CSGB\_SALTI STANDARD; PRT; 151 AA.  
 AC Q827M3;  
 DT 28-FEB-2003 (Rel. 41, Created)  
 DT 28-FEB-2003 (Rel. 41, Last sequence update)  
 DT 10-OCT-2003 (Rel. 42, Last annotation update)  
 DE Minor curlin subunit precursor.  
 GN CSGB OR STY1180 OR T1777.  
 OS *Salmonella typhi*.  
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;  
 OC Enterobacteriaceae; *Salmonella*.  
 OX NCBI\_TaxID=601;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=CT18;  
 RX MEDLINE=21534947; PubMed=11677608;  
 RA Parkhill J., Dougan G., James K.D., Thomson N.R., Pickard D., Wain J.,  
 RA Churcher C., Mungall K.L., Bentley S.D., Holden M.T.G., Sebahia M.,  
 RA Baker S., Basham D., Brooks K., Chillingworth T., Connor P.,  
 RA Cronin A., Davis P., Davies R.M., Dowd L., White N., Farrar J.,  
 RA Feltwell T., Hamlin N., Haque A., Hien T.T., Holroyd S., Jagels K.,  
 RA Krogh A., Larsen T.S., Leather S., Moule S., O'Gaora P., Parry C.,  
 RA Quail M.A., Rutherford K., Simmonds M., Skelton J., Stevens K.,  
 RA Whitehead S., Barrell B.G.;  
 RT "Complete genome sequence of a multiple drug resistant *Salmonella*  
 enterica serovar *Typhi* CT18.";  
 RL Nature 413:848-852(2001).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=Ty2 / ATCC 700931;  
 RX MEDLINE=22531367; PubMed=12644504;  
 RA Deng W., Liou S.-R., Plunkett G. III, Mayhew G.F., Rose D.J.,  
 RA Burland V., Kodyanni V., Schwartz D.C., Blattner F.R.;  
 RT "Comparative genomics of *Salmonella enterica* serovar *Typhi* strains Ty2  
 and CT18.";  
 RL J. Bacteriol. 185:2330-2337(2003).  
 CC -!- FUNCTION: CURLIN IS THE STRUCTURAL SUBUNIT OF THE CURLI. CURLI ARE  
 CC COILED SURFACE STRUCTURES THAT ASSEMBLE PREFERENTIALLY AT GROWTH  
 CC TEMPERATURES BELOW 37 DEGREES CELSIUS. CURLI CAN BIND TO  
 CC FIBRONECTIN. THE MINOR SUBUNIT IS THE NUCLEATION COMPONENT OF  
 CC CURLIN MONOMERS.  
 CC -!- SIMILARITY: BELONGS TO THE CSGA/CSGB FAMILY.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 CC EMBL, X90754; CAA62281.1; -.

CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

DR EMBL; AL627269; CAD08267.1; -;  
DR EMBL; AE016840; AAO69400.1; -;  
KW Fimbrina; Signal; Complete proteome.  
FT SIGNAL 1 21 POTENTIAL.  
FT CHAIN 22 151 MINOR CURLIN SUBUNIT.  
SQ SEQUENCE 151 AA; 16254 MW; 161C543268573495 CRC64;

Query Match 13.7%; Score 106.5; DB 1; Length 151;  
Best Local Similarity 30.7%; Pred. No. 0.015;  
Matches 35; Conservative 17; Mismatches 43; Indels 19; Gaps 5;

QY 55 MAHALQSDARKSE-----TTITQSGYNGADVGQ-GADNST-----IELTQNGFR 98  
Db 18 IATATNYDLARSEYNFAVNELSKSFNQAAIIQGVGTDSARVROGSKLLSVISQEGEN 77  
QY 99 NNATIDOWNAKNSDIT-VGQYGGNNAALVNQTASDSSVMVROVGFNNATANQY 151  
Db 78 NRAKVQDQ--AGNYNFAYIEQTGNANDASISQAYGNSAAIIQKSGNKANITQY 129

## RESULT 6

CSGB\_SALTY STANDARD; PRT; 151 AA.  
AC P55226;  
DT 01-OCT-1996 (Rel. 34, Created)  
DT 01-OCT-1996 (Rel. 34, Last sequence update)  
DT 28-FEB-2003 (Rel. 41, Last annotation update)  
DE Minor curlin subunit precursor (Fimbrin SEF17 minor subunit).  
GN CSGB OR AGFB OR STM143.  
OS Salmonella typhimurium, and  
OS Salmonella enteritidis.  
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;  
OC Enterobacteriaceae; Salmonella.  
OX NCBI\_TaxID=602, 592;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC SPECIES=S.typhimurium; STRAIN=SR-11;  
RX MEDLINE=98117058; PubMed=9457880;  
RA Rømling U., Bian Z., Hammar M., Sierralta W.D., Normark S.;  
RT "Curli" fibers are highly conserved between Salmonella typhimurium and  
RT Escherichia coli with respect to operon structure and regulation.";  
RL J. Bacteriol. 180:722-731(1998).  
RN [2]

RP SEQUENCE FROM N.A.  
RC SPECIES=S.typhimurium; STRAIN=LT2 / SGSC1412 / ATCC 700720;  
RX MEDLINE=21534948; PubMed=11677609;  
RA McClelland M., Sanderson K.E., Spieth J., Clifton S.W., Latreille P.,  
RA Courtney L., Porwollik S., Ali J., Dante M., Du F., Hou S., Layman D.,  
RA Leonard S., Nguyen C., Scott K., Holmes A., Grewal N., Mulvaney E.,  
RA Ryan E., Sun H., Flores L., Miller W., Stoneking T., Nhan M.,  
RA Waterston R., Wilson R.K.;  
RT "Complete genome sequence of Salmonella enterica serovar Typhimurium  
RT LT2.";  
RL Nature 413:852-856(2001).  
RN [3]

RP SEQUENCE FROM N.A.  
RC SPECIES=S.enteritidis; STRAIN=27655-3B;  
RX MEDLINE=96146512; PubMed=8550497;  
RA Collinson S.K., Clouthier S.C., Doran J.L., Banser P.A., Kay W.W.;  
RT "Salmonella enteritidis agfBAC operon encoding thin, aggregative  
RT fimbriae";  
RL J. Bacteriol. 178:662-667(1996).

CC -!- FUNCTION: CURLIN IS THE STRUCTURAL SUBUNIT OF THE CURLI. CURLI ARE  
CC COILED SURFACE STRUCTURES THAT ASSEMBLE PREFERENTIALLY AT GROWTH  
CC TEMPERATURES BELOW 37 DEGREES CELSIUS. CURLI CAN BIND TO  
CC FIBRONECTIN. THE MINOR SUBUNIT IS THE NUCLEATION COMPONENT OF  
CC CURLIN MONOMERS.

CC -!- SIMILARITY: BELONGS TO THE CSGB/CSGB FAMILY.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

DR EMBL; AJ002301; CAA05316.1; -;  
DR EMBL; AE008749; AAL20073.1; -;  
DR EMBL; U43280; AAC43598.1; -;  
DR PIR; JC6040; JC6040.  
KW StyGene; SG10609; csGB.

KW Fimbrina; Signal; Complete proteome.  
FT SIGNAL 1 21 POTENTIAL.  
FT CHAIN 22 151 MINOR CURLIN SUBUNIT.  
SQ SEQUENCE 151 AA; 16182 MW; C0FC5430E6D361D CRC64;

Query Match 13.7%; Score 106.5; DB 1; Length 151;  
Best Local Similarity 30.7%; Pred. No. 0.015;  
Matches 35; Conservative 17; Mismatches 43; Indels 19; Gaps 5;

QY 55 MAHALQSDARKSE-----TTITQSGYNGADVGQ-GADNST-----IELTQNGFR 98  
Db 18 IATATNYDLARSEYNFAVNELSKSFNQAAIIQGVGTDSARVROGSKLLSVISQEGEN 77

QY 99 NNATIDOWNAKNSDIT-VGQYGGNNAALVNQTASDSSVMVROVGFNNATANQY 151  
Db 78 NRAKVQDQ--AGNYNFAYIEQTGNANDASISQAYGNSAAIIQKSGNKANITQY 129

## RESULT 7

YK03 CAEEL STANDARD; PRT; 401 AA.  
AC P34291;  
DT 01-FEB-1994 (Rel. 28, Created)  
DT 30-MAY-2000 (Rel. 39, Last sequence update)  
DT 30-MAY-2000 (Rel. 39, Last annotation update)  
DE Hypothetical 43.5 kDa protein C05B5.3 in chromosome III.  
GN C05B5.3.  
OS Caenorhabditis elegans.

OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;  
OC Rhabditidae; Peloderinae; Caenorhabditis.  
OX NCBI\_TaxID=6239;  
RN [1]  
RP SEQUENCE FROM N.A.

RC STRAIN=Bristol N2;  
RC Wormmore B.J.; 1994) to the EMBL/GenBank/DBJ databases.  
RL Submitted (APR-1994) to the EMBL/GenBank/DBJ databases.  
RN [2]

## REVIEWS.

RA Durbir R.;

RL Submitted (NOV-1998) to the EMBL/GenBank/DBJ databases.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

DR EMBL; Z32679; CAA83596.1; -;

DR PIR; C88571; C88571.

DR WormPep; C05B5.3; C017369.

KW Hypothetical protein.

SQ SEQUENCE 401 AA; 43472 MW; C99DF259EF6C8B55 CRC64;

Query Match 11.8%; Score 92; DB 1; Length 401;

Best Local Similarity 23.2%; Pred. No. 0.8;

Matches 46; Conservative 19; Mismatches 71; Indels 62; Gaps 7;

QY 3 LKVAFAAIVSGSALAGVVPQWGG-----GNHNGGNSGPDYDQVLT 48  
 Db 30 LQYRAVAGSSGCVPAIVPKSGFWENADMIAGLQEQRSQONQNNPQDDPRT 89  
 QY 49 RVVTHE-----MAHALQSDARKSETTI---TQSG--YNGADVGOQA 85  
 Db 90 QSQTGGINGVNPQSSSSNOQPVIIYIARAGSKYKNSVETTFPTNGFNGFGQGNQ 149  
 QY 86 D-----NSTIETQNGFNATIDQWNAKNSDITVGOYGGNNAALVNVQTASD 132  
 Db 150 NTFSSGFFNNQNSQNLQNNFQCN-----QNLGASSGFNNQN---QNSQ 197  
 QY 133 SSVNVRQVGFNNATNQ 150  
 Db 198 NVNGPTSGFNSQTSNQ 215

RESULT 8  
 OMPE\_RICJA STANDARD; PRT; 1656 AA.  
 AC O06653;  
 DT 30-MAY-2000 (Rel. 39, Created)  
 DT 30-MAY-2000 (Rel. 39, Last sequence update)  
 DT 16-OCT-2001 (Rel. 40, Last annotation update)  
 DE Outer membrane protein B precursor (168 kDa surface-layer protein)  
 DE (Surface protein antigen) (Cell surface antigen 5) (ScA5) (rOMP)  
 DE (rOMP B) [Contains: 120 kDa surface-exposed protein (Surface protein  
 antigen) (120 kDa outer membrane protein ompB); 32 kDa beta peptide].  
 GN OMPB.  
 OS Rickettsia japonica.  
 OC Bacteria; Proteobacteria; Alphaproteobacteria; Rickettsiales;  
 OC Rickettsiaceae; Rickettsiae; Rickettsia.  
 OX NCBI\_TaxID=35790;  
 RN SEQUENCE FROM N.A.  
 RC STRAIN=YH;  
 RA Uchiyama T.;  
 RT "Sequencing of the gene encoding the protein rOMP B of Rickettsia  
 japonica";  
 RL Submitted (MAY-1997) to the EMBL/GenBank/DBJ databases.  
 CC -!- FUNCTION: THE 120 kDa SURFACE-EXPOSED PROTEIN IS A MAJOR  
 CC STRUCTURAL PROTEIN WHICH MAY PLAY A ROLE AS A RICKETTSIAL  
 CC VIRULENCE FACTOR AND/OR IMMUNOGEN DURING INFECTION (BY  
 CC similarity).  
 CC -!- FUNCTION: THE 32 kDa BETA PEPTIDE MAY SERVE AS A MEMBRANE ANCHOR  
 CC (BY similarity).  
 CC -!- SUBCELLULAR LOCATION: Cell wall. This bacterium is covered by a S-  
 CC layer with hexagonal symmetry.  
 CC -!- SIMILARITY: BELONGS TO THE RICKETTSIAE OMPA/OMPB FAMILY.  
 CC  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC  
 CC EMBL: AB003681; BAA20138.1;  
 CC InterPro: IPR006315; Autotransport.  
 CC InterPro: IPR005546; Autotransporter.  
 CC Pfam: PF03797; Autotransporter; 1.  
 CC TIGRfams: TIGR01414; autotrans\_bar1; 2.  
 CC Antigen; S-layer; Cell wall.  
 CC CHAIN 1 1338 120 kDa SURFACE-EXPOSED PROTEIN.  
 CC FT CHAIN 1339 1656 32 kDa BETA PEPTIDE.  
 CC FT DOMAIN 528 533 POLY-GLY  
 CC SEQUENCE 1656 AA; 168097 MW; 3132A69C9DD5993F CRC64;  
 Query Match 11.8%; Score 92; DB 1; Length 1656;  
 Best Local Similarity 26.2%; Pred. No. 4;  
 Matches 45; Conservative 19; Mismatches 52; Indels 56; Gaps 9;

QY 6 VAAPAAIVSGSALAGVVPQWGGGNHNGGNSGPDYDQVLTTRVVTHEVAHALQSDARK 65  
 Db 509 VLAAGAITLDSGATI-----TCDIGNGGGAA-----LANDATK 547  
 QY 66 SETTITQSG---YNGADVGOGADNSTIETQNGFNATID- 104  
 Db 548 ---TLTGGANIISANGTINFQANGTIKLTST--QNNIVDCLAIATDQTGVVDASS 602  
 QY 105 QWNAKNSDI--TVQYGGNNAAL-----VNQTASDSSVMVRQVGFNNAT 147  
 Db 603 LTNAQTITISGTIGIGANNNTLQFNIGSKTLLNGNVAINELVIGNNGS 654

RESULT 9  
 GP63\_LEIDO STANDARD; PRT; 590 AA.  
 AC P23223;  
 DT 01-NOV-1991 (Rel. 20, Created)  
 DT 01-NOV-1991 (Rel. 20, Last sequence update)  
 DT 10-OCT-2003 (Rel. 42, Last annotation update)  
 DE Leishmanolysin precursor (EC 3.4.24.36) (Cell surface protease)  
 DE (Major surface glycoprotein) (GP63 protein) (Promastigote surface  
 endopeptidase).  
 GN GP63.  
 OS Leishmania donovani.  
 OC Eukaryota; Euklenozoa; Kinetoplastida; Trypanosomatidae; Leishmania.  
 OX NCBI\_TaxID=5661;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=LV9;  
 RX MEDLINE=92107220; PubMed=1762629;  
 RA Webb J.R., Button L.L., McMaster R.W.;  
 RT "Heterogeneity of the genes encoding the major surface glycoprotein  
 of Leishmania donovani";  
 RL Mol. Biochem. Parasitol. 48:173-184 (1991).  
 CC -!- FUNCTION: Has an integral role during the infection of macrophages  
 CC in the mammalian host.  
 CC -!- CATALYTIC ACTIVITY: Preference for hydrophobic residues at P1 and  
 CC P1' and basic residues at P2 and P3'. A model nonapeptide is  
 CC cleaved at -Ala-Tyr-|-Leu-Lys-Lys-  
 CC -!- COFACTOR: Binds 1 zinc ion per subunit (By similarity).  
 CC -!- SUBCELLULAR LOCATION: Attached to the membrane by a GPI-anchor.  
 CC -!- SIMILARITY: Belongs to peptidase family M8.  
 CC  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC  
 CC EMBL: M60048; AAA29244.1;  
 CC HSSP: P08148; 1LML.  
 CC MEROPS: M08.001.  
 CC InterPro: IPR006025; Pept\_M\_Zn\_BS.  
 CC InterPro: IPR001577; Peptidase\_M8.  
 CC Pfam: PF01457; Peptidase\_M8; 1.  
 CC PRINTS: PR00782; LSHMANOLYSIN.  
 CC PROSITE: PS00142; ZINC PROTEASE; 1.  
 CC Hydrolase; Metalloprotease; Glycoprotein; Metal-binding; Zinc;  
 CC Zymogen; Signal; Cell adhesion; GPI-anchor; Lipoprotein.  
 CC SIGNAL 1 39 POTENTIAL.  
 CC FT PROPEP 40 87 ACTIVATION PEPTIDE.  
 CC FT CHAIN 88 565 LEISHMANOLYSIN.  
 CC FT PROPEP 566 590 REMOVED IN MATURE FORM (BY SIMILARITY).  
 CC FT METAL 251 251 ZINC (CATALYTIC) (BY SIMILARITY).  
 CC FT ACT SITE 252 252 BY SIMILARITY.  
 CC FT METAL 255 255 ZINC (CATALYTIC) (BY SIMILARITY).  
 CC FT METAL 321 321 ZINC (CATALYTIC) (BY SIMILARITY).  
 CC FT DISULFID 112 129 BY SIMILARITY.  
 CC FT DISULFID 178 217 BY SIMILARITY.  
 CC FT DISULFID 301 373 BY SIMILARITY.



```
FT DISULFID 380 443 BY SIMILARITY.
FT DISULFID 393 412 BY SIMILARITY.
FT DISULFID 402 477 BY SIMILARITY.
FT DISULFID 454 498 BY SIMILARITY.
FT DISULFID 503 553 BY SIMILARITY.
FT DISULFID 523 546 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 287 287 GPI-anchor amidated asparagine (By
FT LIPID 565 similarity).
FT SEQUENCE 590 AA; 62950 MW; 0FB315D299659F58 CRC64;
SQ
Query Match 11.7%; Score 91.5; DB 1; Length 590;
Best Local Similarity 30.5%; Pred. No. 1.4;
Matches 43; Conservative 12; Mismatches 39; Indels 47; Gaps 9;
QY 43 YDQLVTRVVTHEMAHALQ-----SDAR-----KSETITQTS--GYG 76
DB 241 YDQLVTRVVTHEMAHALGFSVVFFRDARILESINVRHXDFDPVINSSTAVAKAREQVG 300
QY 77 NGA-----DVGGQADNSTIETQNGFRNNATIDQWNAKNSDITVGQYGGNNAALVQ 128
DB 301 CGTLEYLEWEDQGGAGSASHKX-----RNAQ-DELMAPASD--AGYYSALTWAIQD 351
QY 129 TA----SDSSVMVRQVGFQGNNA 146
DB 352 LGFYQADFS-KAEEMPWGRNA 371
RESULT 10
PER DROWI STANDARD; PRT; 1093 AA.
AC Q03297; O18421; O18422; P91721; P91722;
DT 01-OCT-1993 (Rel. 27, Created)
DT 15-JUL-1998 (Rel. 36, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Period circadian protein (Fragment).
GN PER.
OS Drosophila willistoni (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
CX NCBI_TaxID=7260;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Various strains;
RX MEDLINE=93196482; PubMed=8450754;
RA Peixoto A.A., Campesan S., Costa R.H., Kyriacou C.P.;
RT "Molecular evolution of a repetitive region within the per gene of
RL Drosophila.";
RN [2]
RP SEQUENCE OF 579-646 FROM N.A.
RX MEDLINE=93196482; PubMed=8450754;
RA Peixoto A.A., Campesan S., Costa R.H., Kyriacou C.P.;
RT "Molecular evolution of a repetitive region within the per gene of
RL Drosophila.";
RN [3]
RP FUNCTION: Essential for biological clock functions. Determines the
RX period length of circadian and ultradian rhythms; an increase in
CC PER dosage leads to shortened circadian rhythms and a decrease
CC leads to lengthened circadian rhythms. Essential for the circadian
CC rhythmicity of locomotor activity, eclosion behavior, and for the
CC rhythmic component of the male courtship song that originates in
CC the thoracic nervous system. The biological cycle depends on the
CC rhythmic formation and nuclear localization of the TIM-PER
CC complex. Light induces the degradation of TIM, which promotes
CC elimination of PER. Nuclear activity of the heterodimer
CC coordinately regulates PER and TIM transcription through a
CC negative feedback loop. Behaves as a negative element in circadian
CC transcriptional loop. Does not appear to bind DNA, suggesting
CC indirect transcriptional inhibition (By similarity).
CC -!- SUBUNIT: Forms heterodimer with timeless (TIM); the complex then
CC translocates into the nucleus (By similarity).
```

```
CC -!- SUBCELLULAR LOCATION: Nuclear at specific periods of the day.
CC First accumulates in the perinuclear region about one hour before
CC translocation into the nucleus. Interaction with Tim is required
CC for nuclear localization (By similarity).
CC -!- PTM: PHOSPHORYLATED WITH A CIRCADIAN RHYTHMICITY, PROBABLY BY THE
CC DOUBLE-TIME PROTEIN (DBT). PHOSPHORYLATION COULD BE IMPLICATED IN
CC THE STABILITY OF PER MONOMER AND IN THE FORMATION OF HETERODIMER
CC PER-TIM (BY SIMILARITY).
CC -!- MISCELLANEOUS: THE SEQUENCE SHOWN IS THAT OF STRAIN AT1XCO.
CC -!- SIMILARITY: Contains 2 PAS (PER-ARNT-SIM) dimerization domains.
CC -!- SIMILARITY: Contains 1 PAS-associated C-terminal (PAC) domain.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; U51055; AAB41360.1; -
CC EMBL; U51056; AAB41361.1; -
CC EMBL; U51057; AAB41362.1; -
CC EMBL; U51058; AAB41363.1; -
CC EMBL; U51059; AAB41364.1; -
CC EMBL; U51060; AAB41365.1; -
CC EMBL; U51061; AAB41366.1; -
CC EMBL; U51062; AAB41367.1; -
CC EMBL; U51063; AAB41368.1; -
CC EMBL; U51064; AAB41369.1; -
CC EMBL; U51065; AAB41370.1; -
CC EMBL; U51066; AAB41371.1; -
CC EMBL; U51067; AAB41372.1; -
CC EMBL; U51068; AAB41373.1; -
CC EMBL; U51069; AAB41374.1; -
CC EMBL; U51070; AAB41375.1; -
CC EMBL; U51071; AAB41376.1; -
CC EMBL; U51072; AAB41377.1; -
CC EMBL; U51073; AAB41378.1; -
CC EMBL; U51074; AAB41379.1; -
CC EMBL; U51075; AAB41380.1; -
CC EMBL; U51076; AAB41381.1; -
CC EMBL; U51077; AAB41382.1; -
CC EMBL; U51078; AAB41383.1; -
CC EMBL; U51079; AAB41384.1; -
CC EMBL; U51080; AAB41385.1; -
CC EMBL; U51081; AAB41386.1; -
CC EMBL; U51082; AAB41387.1; -
CC EMBL; U51083; AAB41388.1; -
CC EMBL; U51084; AAB41389.1; -
CC EMBL; U51085; AAB41390.1; -
CC EMBL; U51086; AAB41391.1; -
CC EMBL; U51087; AAB41392.1; -
CC EMBL; U51088; AAB41393.1; -
CC EMBL; U51089; AAB41394.1; -
CC EMBL; U51090; AAB41395.1; -
CC EMBL; U51091; AAB41396.1; -
CC EMBL; U51092; AAB41397.1; -
CC EMBL; U51093; AAB41398.1; -
CC EMBL; U51094; AAB41399.1; -
CC EMBL; U51095; AAB41400.1; -
CC EMBL; U51096; AAB41401.1; -
CC EMBL; U51097; AAB41402.1; -
CC EMBL; U51098; AAB41403.1; -
CC EMBL; U51099; AAB41404.1; -
CC EMBL; U51100; AAB41405.1; -
CC EMBL; U51101; AAB41406.1; -
CC EMBL; U51102; AAB41407.1; -
CC EMBL; U51103; AAB41408.1; -
CC EMBL; U51104; AAB41409.1; -
CC EMBL; U51105; AAB41410.1; -
CC EMBL; U51106; AAB41411.1; -
CC EMBL; U51107; AAB41412.1; -
CC EMBL; U51108; AAB41413.1; -
CC EMBL; U51109; AAB41414.1; -
CC EMBL; U51110; AAB41415.1; -
CC EMBL; U51111; AAB41416.1; -
CC EMBL; U51112; AAB41417.1; -
CC EMBL; U51113; AAB41418.1; -
CC EMBL; U51114; AAB41419.1; -
CC EMBL; U51115; AAB41420.1; -
CC EMBL; U51116; AAB41421.1; -
CC EMBL; U51117; AAB41422.1; -
CC EMBL; U51118; AAB41423.1; -
CC EMBL; U51119; AAB41424.1; -
CC EMBL; U51120; AAB41425.1; -
CC EMBL; U51121; AAB41426.1; -
CC EMBL; U51122; AAB41427.1; -
CC EMBL; U51123; AAB41428.1; -
CC EMBL; U51124; AAB41429.1; -
CC EMBL; U51125; AAB41430.1; -
CC EMBL; U51126; AAB41431.1; -
CC EMBL; U51127; AAB41432.1; -
CC EMBL; U51128; AAB41433.1; -
CC EMBL; U51129; AAB41434.1; -
CC EMBL; U51130; AAB41435.1; -
CC EMBL; U51131; AAB41436.1; -
CC EMBL; U51132; AAB41437.1; -
CC EMBL; U51133; AAB41438.1; -
CC EMBL; U51134; AAB41439.1; -
CC EMBL; U51135; AAB41440.1; -
CC EMBL; U51136; AAB41441.1; -
CC EMBL; U51137; AAB41442.1; -
CC EMBL; U51138; AAB41443.1; -
CC EMBL; U51139; AAB41444.1; -
CC EMBL; U51140; AAB41445.1; -
CC EMBL; U51141; AAB41446.1; -
CC EMBL; U51142; AAB41447.1; -
CC EMBL; U51143; AAB41448.1; -
CC EMBL; U51144; AAB41449.1; -
CC EMBL; U51145; AAB41450.1; -
CC EMBL; U51146; AAB41451.1; -
CC EMBL; U51147; AAB41452.1; -
CC EMBL; U51148; AAB41453.1; -
CC EMBL; U51149; AAB41454.1; -
CC EMBL; U51150; AAB41455.1; -
CC EMBL; U51151; AAB41456.1; -
CC EMBL; U51152; AAB41457.1; -
CC EMBL; U51153; AAB41458.1; -
CC EMBL; U51154; AAB41459.1; -
CC EMBL; U51155; AAB41460.1; -
CC EMBL; U51156; AAB41461.1; -
CC EMBL; U51157; AAB41462.1; -
CC EMBL; U51158; AAB41463.1; -
CC EMBL; U51159; AAB41464.1; -
CC EMBL; U51160; AAB41465.1; -
CC EMBL; U51161; AAB41466.1; -
CC EMBL; U51162; AAB41467.1; -
CC EMBL; U51163; AAB41468.1; -
CC EMBL; U51164; AAB41469.1; -
CC EMBL; U51165; AAB41470.1; -
CC EMBL; U51166; AAB41471.1; -
CC EMBL; U51167; AAB41472.1; -
CC EMBL; U51168; AAB41473.1; -
CC EMBL; U51169; AAB41474.1; -
CC EMBL; U51170; AAB41475.1; -
CC EMBL; U51171; AAB41476.1; -
CC EMBL; U51172; AAB41477.1; -
CC EMBL; U51173; AAB41478.1; -
CC EMBL; U51174; AAB41479.1; -
CC EMBL; U51175; AAB41480.1; -
CC EMBL; U51176; AAB41481.1; -
CC EMBL; U51177; AAB41482.1; -
CC EMBL; U51178; AAB41483.1; -
CC EMBL; U51179; AAB41484.1; -
CC EMBL; U51180; AAB41485.1; -
CC EMBL; U51181; AAB41486.1; -
CC EMBL; U51182; AAB41487.1; -
CC EMBL; U51183; AAB41488.1; -
CC EMBL; U51184; AAB41489.1; -
CC EMBL; U51185; AAB41490.1; -
CC EMBL; U51186; AAB41491.1; -
CC EMBL; U51187; AAB41492.1; -
CC EMBL; U51188; AAB41493.1; -
CC EMBL; U51189; AAB41494.1; -
CC EMBL; U51190; AAB41495.1; -
CC EMBL; U51191; AAB41496.1; -
CC EMBL; U51192; AAB41497.1; -
CC EMBL; U51193; AAB41498.1; -
CC EMBL; U51194; AAB41499.1; -
CC EMBL; U51195; AAB41500.1; -
CC EMBL; U51196; AAB41501.1; -
CC EMBL; U51197; AAB41502.1; -
CC EMBL; U51198; AAB41503.1; -
CC EMBL; U51199; AAB41504.1; -
CC EMBL; U51200; AAB41505.1; -
CC EMBL; U51201; AAB41506.1; -
CC EMBL; U51202; AAB41507.1; -
CC EMBL; U51203; AAB41508.1; -
CC EMBL; U51204; AAB41509.1; -
CC EMBL; U51205; AAB41510.1; -
CC EMBL; U51206; AAB41511.1; -
CC EMBL; U51207; AAB41512.1; -
CC EMBL; U51208; AAB41513.1; -
CC EMBL; U51209; AAB41514.1; -
CC EMBL; U51210; AAB41515.1; -
CC EMBL; U51211; AAB41516.1; -
CC EMBL; U51212; AAB41517.1; -
CC EMBL; U51213; AAB41518.1; -
CC EMBL; U51214; AAB41519.1; -
CC EMBL; U51215; AAB41520.1; -
CC EMBL; U51216; AAB41521.1; -
CC EMBL; U51217; AAB41522.1; -
CC EMBL; U51218; AAB41523.1; -
CC EMBL; U51219; AAB41524.1; -
CC EMBL; U51220; AAB41525.1; -
CC EMBL; U51221; AAB41526.1; -
CC EMBL; U51222; AAB41527.1; -
CC EMBL; U51223; AAB41528.1; -
CC EMBL; U51224; AAB41529.1; -
CC EMBL; U51225; AAB41530.1; -
CC EMBL; U51226; AAB41531.1; -
CC EMBL; U51227; AAB41532.1; -
CC EMBL; U51228; AAB41533.1; -
CC EMBL; U51229; AAB41534.1; -
CC EMBL; U51230; AAB41535.1; -
CC EMBL; U51231; AAB41536.1; -
CC EMBL; U51232; AAB41537.1; -
CC EMBL; U51233; AAB41538.1; -
CC EMBL; U51234; AAB41539.1; -
CC EMBL; U51235; AAB41540.1; -
CC EMBL; U51236; AAB41541.1; -
CC EMBL; U51237; AAB41542.1; -
CC EMBL; U51238; AAB41543.1; -
CC EMBL; U51239; AAB41544.1; -
CC EMBL; U51240; AAB41545.1; -
CC EMBL; U51241; AAB41546.1; -
CC EMBL; U51242; AAB41547.1; -
CC EMBL; U51243; AAB41548.1; -
CC EMBL; U51244; AAB41549.1; -
CC EMBL; U51245; AAB41550.1; -
CC EMBL; U51246; AAB41551.1; -
CC EMBL; U51247; AAB41552.1; -
CC EMBL; U51248; AAB41553.1; -
CC EMBL; U51249; AAB41554.1; -
CC EMBL; U51250; AAB41555.1; -
CC EMBL; U51251; AAB41556.1; -
CC EMBL; U51252; AAB41557.1; -
CC EMBL; U51253; AAB41558.1; -
CC EMBL; U51254; AAB41559.1; -
CC EMBL; U51255; AAB41560.1; -
CC EMBL; U51256; AAB41561.1; -
CC EMBL; U51257; AAB41562.1; -
CC EMBL; U51258; AAB41563.1; -
CC EMBL; U51259; AAB41564.1; -
CC EMBL; U51260; AAB41565.1; -
CC EMBL; U51261; AAB41566.1; -
CC EMBL; U51262; AAB41567.1; -
CC EMBL; U51263; AAB41568.1; -
CC EMBL; U51264; AAB41569.1; -
CC EMBL; U51265; AAB41570.1; -
CC EMBL; U51266; AAB41571.1; -
CC EMBL; U51267; AAB41572.1; -
CC EMBL; U51268; AAB41573.1; -
CC EMBL; U51269; AAB41574.1; -
CC EMBL; U51270; AAB41575.1; -
CC EMBL; U51271; AAB41576.1; -
CC EMBL; U51272; AAB41577.1; -
CC EMBL; U51273; AAB41578.1; -
CC EMBL; U51274; AAB41579.1; -
CC EMBL; U51275; AAB41580.1; -
CC EMBL; U51276; AAB41581.1; -
CC EMBL; U51277; AAB41582.1; -
CC EMBL; U51278; AAB41583.1; -
CC EMBL; U51279; AAB41584.1; -
CC EMBL; U51280; AAB41585.1; -
CC EMBL; U51281; AAB41586.1; -
CC EMBL; U51282; AAB41587.1; -
CC EMBL; U51283; AAB41588.1; -
CC EMBL; U51284; AAB41589.1; -
CC EMBL; U51285; AAB41590.1; -
CC EMBL; U51286; AAB41591.1; -
CC EMBL; U51287; AAB41592.1; -
CC EMBL; U51288; AAB41593.1; -
CC EMBL; U51289; AAB41594.1; -
CC EMBL; U51290; AAB41595.1; -
CC EMBL; U51291; AAB41596.1; -
CC EMBL; U51292; AAB41597.1; -
CC EMBL; U51293; AAB41598.1; -
CC EMBL; U51294; AAB41599.1; -
CC EMBL; U51295; AAB41600.1; -
CC EMBL; U51296; AAB41601.1; -
CC EMBL; U51297; AAB41602.1; -
CC EMBL; U51298; AAB41603.1; -
CC EMBL; U51299; AAB41604.1; -
CC EMBL; U51300; AAB41605.1; -
CC EMBL; U51301; AAB41606.1; -
CC EMBL; U51302; AAB41607.1; -
CC EMBL; U51303; AAB41608.1; -
CC EMBL; U51304; AAB41609.1; -
CC EMBL; U51305; AAB41610.1; -
CC EMBL; U51306; AAB41611.1; -
CC EMBL; U51307; AAB41612.1; -
CC EMBL; U51308; AAB41613.1; -
CC EMBL; U51309; AAB41614.1; -
CC EMBL; U51310; AAB41615.1; -
CC EMBL; U51311; AAB41616.1; -
CC EMBL; U51312; AAB41617.1; -
CC EMBL; U51313; AAB41618.1; -
CC EMBL; U51314; AAB41619.1; -
CC EMBL; U51315; AAB41620.1; -
CC EMBL; U51316; AAB41621.1; -
CC EMBL; U51317; AAB41622.1; -
CC EMBL; U51318; AAB41623.1; -
CC EMBL; U51319; AAB41624.1; -
CC EMBL; U51320; AAB41625.1; -
CC EMBL; U51321; AAB41626.1; -
CC EMBL; U51322; AAB41627.1; -
CC EMBL; U51323; AAB41628.1; -
CC EMBL; U51324; AAB41629.1; -
CC EMBL; U51325; AAB41630.1; -
CC EMBL; U51326; AAB41631.1; -
CC EMBL; U51327; AAB41632.1; -
CC EMBL; U51328; AAB41633.1; -
CC EMBL; U51329; AAB41634.1; -
CC EMBL; U51330; AAB41635.1; -
CC EMBL; U51331; AAB41636.1; -
CC EMBL; U51332; AAB41637.1; -
CC EMBL; U51333; AAB41638.1; -
CC EMBL; U51334; AAB41639.1; -
CC EMBL; U51335; AAB41640.1; -
CC EMBL; U51336; AAB41641.1; -
CC EMBL; U51337; AAB41642.1; -
CC EMBL; U51338; AAB41643.1; -
CC EMBL; U51339; AAB41644.1; -
CC EMBL; U51340; AAB41645.1; -
CC EMBL; U51341; AAB41646.1; -
CC EMBL; U51342; AAB41647.1; -
CC EMBL; U51343; AAB41648.1; -
CC EMBL; U51344; AAB41649.1; -
CC EMBL; U51345; AAB41650.1; -
CC EMBL; U51346; AAB41651.1; -
CC EMBL; U51347; AAB41652.1; -
CC EMBL; U51348; AAB41653.1; -
CC EMBL; U51349; AAB41654.1; -
CC EMBL; U51350; AAB41655.1; -
CC EMBL; U51351; AAB41656.1; -
CC EMBL; U51352; AAB41657.1; -
CC EMBL; U51353; AAB41658.1; -
CC EMBL; U51354; AAB41659.1; -
CC EMBL; U51355; AAB41660.1; -
CC EMBL; U51356; AAB41661.1; -
CC EMBL; U51357; AAB41662.1; -
CC EMBL; U51358; AAB41663.1; -
CC EMBL; U51359; AAB41664.1; -
CC EMBL; U51360; AAB41665.1; -
CC EMBL; U51361; AAB41666.1; -
CC EMBL; U51362; AAB41667.1; -
CC EMBL; U51363; AAB41668.1; -
CC EMBL; U51364; AAB41669.1; -
CC EMBL; U51365; AAB41670.1; -
CC EMBL; U51366; AAB41671.1; -
CC EMBL; U51367; AAB41672.1; -
CC EMBL; U51368; AAB41673.1; -
CC EMBL; U51369; AAB41674.1; -
CC EMBL; U51370; AAB41675.1; -
CC EMBL; U51371; AAB41676.1; -
CC EMBL; U51372; AAB41677.1; -
CC EMBL; U51373; AAB41678.1; -
CC EMBL; U51374; AAB41679.1; -
CC EMBL; U51375; AAB41680.1; -
CC EMBL; U51376; AAB41681.1; -
CC EMBL; U51377; AAB41682.1; -
CC EMBL; U51378; AAB41683.1; -
CC EMBL; U51379; AAB41684.1; -
CC EMBL; U51380; AAB41685.1; -
CC EMBL; U51381; AAB41686.1; -
CC EMBL; U51382; AAB41687.1; -
CC EMBL; U51383; AAB41688.1; -
CC EMBL; U51384; AAB41689.1; -
CC EMBL; U51385; AAB41690.1; -
CC EMBL; U51386; AAB41691.1; -
CC EMBL; U51387; AAB41692.1; -
CC EMBL; U51388; AAB41693.1; -
CC EMBL; U51389; AAB41694.1; -
CC EMBL; U51390; AAB41695.1; -
CC EMBL; U51391; AAB41696.1; -
CC EMBL; U51392; AAB41697.1; -
CC EMBL; U51393; AAB41698.1; -
CC EMBL; U51394; AAB41699.1; -
CC EMBL; U51395; AAB41700.1; -
CC EMBL; U51396; AAB41701.1; -
CC EMBL; U51397; AAB41702.1; -
CC EMBL; U51398; AAB41703.1; -
CC EMBL; U51399; AAB41704.1; -
CC EMBL; U51400; AAB41705.1; -
CC EMBL; U51401; AAB41706.1; -
CC EMBL; U51402; AAB41707.1; -
CC EMBL; U51403; AAB41708.1; -
CC EMBL; U51404; AAB41709.1; -
CC EMBL; U51405; AAB41710.1; -
CC EMBL; U51406; AAB41711.1; -
CC EMBL; U51407; AAB41712.1; -
CC EMBL; U51408; AAB41713.1; -
CC EMBL; U51409; AAB41714.1; -
CC EMBL; U51410; AAB41715.1; -
CC EMBL; U51411; AAB41716.1; -
CC EMBL; U51412; AAB41717.1; -
CC EMBL; U51413; AAB41718.1; -
CC EMBL; U51414; AAB41719.1; -
CC EMBL; U51415; AAB41720.1; -
CC EMBL; U51416; AAB41721.1; -
CC EMBL; U51417; AAB41722.1; -
CC EMBL; U51418; AAB41723.1; -
CC EMBL; U51419; AAB41724.1; -
CC EMBL; U51420; AAB41725.1; -
CC EMBL; U51421; AAB41726.1; -
CC EMBL; U51422; AAB41727.1; -
CC EMBL; U51423; AAB41728.1; -
CC EMBL; U51424; AAB41729.1; -
CC EMBL; U51425; AAB41730.1; -
CC EMBL; U51426; AAB41731.1; -
CC EMBL; U51427; AAB41732.1; -
CC EMBL; U51428; AAB41733.1; -
CC EMBL; U51429; AAB41734.1; -
CC EMBL; U51430; AAB41735.1; -
CC EMBL; U51431; AAB41736.1; -
CC EMBL; U51432; AAB41737.1; -
CC EMBL; U51433; AAB41738.1; -
CC EMBL; U51434; AAB41739.1; -
CC EMBL; U51435; AAB41740.1; -
CC EMBL; U51436; AAB41741.1; -
CC EMBL; U51437; AAB41742.1; -
CC EMBL; U51438; AAB41743.1; -
CC EMBL; U51439; AAB41744.1; -
CC EMBL; U51440; AAB41745.1; -
CC EMBL; U51441; AAB41746.1; -
CC EMBL; U51442; AAB41747.1; -
CC EMBL; U51443; AAB41748.1; -
CC EMBL; U51444; AAB41749.1; -
CC EMBL; U51445; AAB41750.1; -
CC EMBL; U51446; AAB41751.1; -
CC EMBL; U51447; AAB41752.1; -
CC EMBL; U51448; AAB41753.1; -
CC EMBL; U51449; AAB41754.1; -
CC EMBL; U51450; AAB41755.1; -
CC EMBL; U51451; AAB41756.1; -
CC EMBL; U51452; AAB41757.1; -
CC EMBL; U51453; AAB41758.1; -
CC EMBL; U51454; AAB41759.1; -
CC EMBL; U51455; AAB41760.1; -
CC EMBL; U51456; AAB41761.1; -
CC EMBL; U51457; AAB41762.1; -
CC EMBL; U51458; AAB41763.1; -
CC EMBL; U51459; AAB41764.1; -
CC EMBL; U51460; AAB41765.1; -
CC EMBL; U51461; AAB41766.1; -
CC EMBL; U51462; AAB41767.1; -
CC EMBL; U51463; AAB41768.1; -
CC EMBL; U51464; AAB41769.1; -
CC EMBL; U51465; AAB41770.1; -
CC EMBL; U51466; AAB41771.1; -
CC EMBL; U51467; AAB41772.1; -
CC EMBL; U51468; AAB41773.1; -
CC EMBL; U51469; AAB41774.1; -
CC EMBL; U51470; AAB41775.1; -
CC EMBL; U51471; AAB41776.1; -
CC EMBL; U51472; AAB41777.1; -
CC EMBL; U51473; AAB41778.1; -
CC EMBL; U51474; AAB41779.1; -
CC EMBL; U51475; AAB41780.1; -
CC EMBL; U51476; AAB41781.1; -
CC EMBL; U51477; AAB41782.1; -
CC EMBL; U51478; AAB41783.1; -
CC EMBL; U51479; AAB41784.1; -
CC EMBL; U51480; AAB41785.1; -
CC EMBL; U51481; AAB41786.1; -
CC EMBL; U51482; AAB41787.1; -
CC EMBL; U51483; AAB41788.1; -
CC EMBL; U51484; AAB41789.1; -
CC EMBL; U51485; AAB41790.1; -
CC EMBL; U51486; AAB41791.1; -
CC EMBL; U51487; AAB41792.1; -
CC EMBL; U51488; AAB41793.1; -
CC EMBL; U51489; AAB41794.1; -
CC EMBL; U51490; AAB41795.1; -
CC EMBL; U51491; AAB41796.1; -
CC EMBL; U51492; AAB41797.1; -
CC EMBL; U51493; AAB41798.1; -
CC EMBL; U51494; AAB41799.1; -
CC EMBL; U51495; AAB41800.1; -
CC EMBL; U51496; AAB41801.1; -
CC EMBL; U51497; AAB41802.1; -
CC EMBL; U51498; AAB41803.1; -
CC EMBL; U51499; AAB41804.1; -
CC EMBL; U51500; AAB41805.1; -
CC EMBL; U51501; AAB41806.1; -
CC EMBL; U51
```

```

FT VARIANT 734 734 MISSING (IN STRAINS MANAUS 4, PORTO
FT ALEGRE 1 AND PORTO ALEGRE 2).
FT VARIANT 747 747 S -> A (IN STRAINS GUADELOUPE AND GUANA).
FT VARIANT 764 766 MISSING (IN STRAINS MANAUS 3).
FT VARIANT 886 886 A -> T (IN STRAIN 0811.4).
FT NON TER 1093
SQ SEQUENCE 1093 AA, 115896 MW; AB6DE050267EC187 CRC64;

Query Match 11.7%; Score 91; DB 1; Length 1093;
Best Local Similarity 26.1%; Pred.No. 3.1;
Matches 24; Conservative 8; Mismatches 46; Indels 14; Gaps 2;

QY 27 GGGGNGHGGGSGPDYDQLVTRVTHMAHALQSDARKSETTITQSGYGNAGDVGGQAD 86
DB 719 GGGGGGGGGGGGGGGGLPLFL-----DVTHSSSSQNGKPTGVAAGGAGGGVGGGG-- 770

QY 87 NSTEITQNGFRNRNATIDOWNAKNSDITVQGY 118
DB 771 -----SCSLGCGNGVNGSGNGNSQPTNQY 796

RESULT 11
GP63_LEIMA STANDARD; PRT; 602 AA.
AC P08148; P15906;
DT 01-AUG-1988 (Rel. 08, Created)
DT 01-APR-1990 (Rel. 14, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Leishmanolysin precursor (EC 3.4.24.36) (Cell surface protease)
DE (Major surface glycoprotein) (GP63 protein) (Promastigote surface
DE endopeptidase).
GN GP63.
OS Leishmania major.
OC Eukaryota; Eukaryota; Kinetoplastida; Trypanosomatidae; Leishmania.
OX NCBI_TaxID:5664;
[1]
SEQUENCE FROM N.A., AND SEQUENCE OF 101-123.
RX MEDLINE=88154764; PubMed=3346625;
RA Button L.L., McMaster W.R.;
RT "Molecular cloning of the major surface antigen of leishmania.";
RL J. Exp. Med. 167:724-729 (1988).
[2]
REVISIONS.
RA Button L.L., McMaster W.R.;
RL J. Exp. Med. 171:589-589 (1990).
[3]
GPI-ANCHOR.
RX MEDLINE=91009116; PubMed=2145267;
RA Schneider P., Ferguson M.A.J., McConville M.J., Mehlett A.,
RA Homans S.W., Bordier C.;
RT "Structure of the glycosyl-phosphatidylinositol membrane anchor of
RT the Leishmania major promastigote surface protease.";
RL J. Biol. Chem. 265:16955-16964 (1990).
[4]
X-RAY CRYSTALLOGRAPHY (2.6 ANGSTROMS).
RX MEDLINE=95406217; PubMed=7675789;
RA Schlagenhauf E., Stges R., Metcalf P.;
RT "Crystallization and preliminary X-ray diffraction studies of
RT Leishmanolysin, the major surface metalloproteinase from Leishmania
RT major.";
RL Proteins 22:58-66 (1995).
[5]
X-RAY CRYSTALLOGRAPHY (1.86 ANGSTROMS).
RX MEDLINE=98416698; PubMed=9739094;
RA Schlagenhauf E., Stges R., Metcalf P.;
RT "The crystal structure of the Leishmania major surface proteinase
RT Leishmanolysin.";
RL Structure 6:1035-1046 (1998).
CC -1- FUNCTION: Has an integral role during the infection of macrophages
CC in the mammalian host.
CC -1- CATALYTIC ACTIVITY: Preference for hydrophobic residues at P1 and
CC P1', and basic residues at P2 and P3'. A model nonapeptide is
CC cleaved at -Ala-Iyr-I-Leu-Lys-Lys-.

```

```

CC -1- COPACTOR: Binds 1 zinc ion per subunit.
CC -1- SUBCELLULAR LOCATION: Attached to the membrane by a GPI-anchor.
CC -1- PTM: THE PHOSPHATIDYLINOSITOL MOIETY OF THE GPI-ANCHOR CONTAINS A
CC FULLY SATURATED, UNBRANCHED 1-O-ALKYL CHAIN (MAINLY C24:0) AND A
CC MIXTURE OF FULLY SATURATED UNBRANCHED 2-O-ACYL CHAINS (C12:0,
CC C14:0, C16:0, AND C18:0).
CC -1- SIMILARITY: Belongs to peptidase family M8.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; Y00647; CAA68673.1; --
CC PIR; P02221; P02221.
CC PDB; 1LMG; 17-SEP-97.
CC MEROPS; M08.001; --
CC InterPro; IPR006025; Pept M Zn BS.
CC InterPro; IPR001577; Peptidase_M8.
CC Pfam; PF01457; Peptidase_M8_1.
CC PRINTS; PR00782; LSHMANOLYSIN.
CC PROSITE; PS00142; ZINC_PROTEASE; 1.
CC Hydrolase; Metalloprotease; Glycoprotein; Metal-binding; Zinc;
CC Zymogen; Signal; Cell adhesion; GPI-anchor; 3D-structure; Lipoprotein.
CC SIGNAL 1 39 POTENTIAL.
CC PROPEP 40 100 ACTIVATION PEPTIDE.
CC CHAIN 101 577 LEISHMANOLYSIN.
CC PROPEP 578 602 REMOVED IN MATURE FORM.
CC METAL 264 265 ZINC (CATALYTIC).
CC ACT SITE 268 268
CC METAL 334 334 ZINC (CATALYTIC).
CC METAL 125 142
CC DISULFID 191 230
CC DISULFID 314 386
CC DISULFID 393 455
CC DISULFID 406 425
CC DISULFID 415 489
CC DISULFID 466 510
CC DISULFID 515 565
CC DISULFID 535 558
CC CARBOHYD 300 300
CC CARBOHYD 407 407
CC LIPID 577 577
CC STRAND 101 102
CC STRAND 107 108
CC STRAND 111 114
CC STRAND 116 119
CC TURN 121 122
CC TURN 128 129
CC STRAND 131 133
CC STRAND 139 141
CC STRAND 144 146
CC HELIX 150 158
CC TURN 159 159
CC TURN 160 169
CC TURN 170 171
CC STRAND 172 174
CC STRAND 177 178
CC STRAND 180 181
CC TURN 189 190
CC TURN 191 193
CC HELIX 198 202
CC TURN 203 203
CC STRAND 205 206
CC STRAND 210 215
CC TURN 221 222
CC STRAND 226 232
CC TURN 234 235
CC STRAND 238 244
CC
CC N-LINKED (GLCNAC. . .) (POTENTIAL).
CC N-LINKED (GLCNAC. . .) (POTENTIAL).
CC GPI-anchor amidated asparagine.

```

```

314 CDTLEYLEVEDQGAGSAGSHKM 337

Db
RESULT 12
GLRL_HUMAN
ID GLRL_HUMAN STANDARD; PET; 906 AA.
AC P42261;
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 15-MAR-2004 (Rel. 43, Last annotation update)
DE Glutamate receptor 1 precursor (GluR-1) (GluR-A) (GluR-K1) (Glutamate
GN GRIAL OR GLUR1 OR GLUH1.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=92159065; PubMed=1311100;
RA Sun W., Ferrer-Montiel A.V., Schinder A.F., McPherson J.P.,
RA Evans G.A., Montal M.;
RT "Molecular cloning, chromosomal mapping, and functional expression of
RT human brain glutamate receptors."
RL Proc. Natl. Acad. Sci. U.S.A. 89:1443-1447(1992).
RN [2]
RP SEQUENCE FROM N.A.
RX TISSUE=Hippocampus;
RX MEDLINE=92329975; PubMed=1320959;
RA Potier M.C., Spillantini M.G., Carter N.P.;
RT "The human glutamate receptor cDNA GluR1: cloning, sequencing,
RT expression and localization to chromosome 5."
RL DNA Seq. 2:211-218(1992).
RN [3]
RP SEQUENCE FROM N.A.
RX MEDLINE=91352027; PubMed=1652753;
RA Puckett C., Gomez C.M., Korenberg J.R., Tung H., Meier T.J.,
RA Chen X.N., Hood L.E.;
RT "Molecular cloning and chromosomal localization of one of the human
RT glutamate receptor genes."
RL Proc. Natl. Acad. Sci. U.S.A. 88:7557-7561(1991).
CC -!- FUNCTION: L-glutamate acts as an excitatory neurotransmitter at
CC many synapses in the central nervous system. The postsynaptic
CC actions of glu are mediated by a variety of receptors that are
CC named according to their selective agonists.
CC -!- SUBCELLULAR LOCATION: Integral membrane protein.
CC -!- ALTERNATIVE PRODUCTS:
CC Event=Alternative splicing; Named isoforms=2;
CC Name=Flp;
CC IsoId=P42261-1; Sequence=Displayed;
CC Name=Flp;
CC IsoId=P42261-2; Sequence=VSP_000092, VSP_000093, VSP_000094,
CC VSP_000095, VSP_000096;
CC -!- TISSUE SPECIFICITY: Widely expressed in brain.
CC -!- MISCELLANEOUS: This receptor binds AMPA (quisqualate) > glutamate >
CC kainate.
CC -!- SIMILARITY: Belongs to the ligand-gated ionic channel family.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch)
CC
CC EMBL; M81886; AAA58395.1; -
CC EMBL; X58633; CAA41491.1; -
CC EMBL; M64752; AAA58613.1; -
CC PIR; A40222; A40222.
CC PIR; S25852; S25852.
CC HSSP; P19491; 1GR2.
CC Genew; HGNC:4571; GRIAL.

```

```

247 249
256 269
270 271
274 279
280 281
283 286
289 291
296 299
302 312
313 313
315 316
320 322
328 332
335 335
337 339
341 342
343 343
344 345
353 353
356 364
365 366
369 370
372 374
380 383
386 390
394 395
396 397
398 399
402 404
413 414
417 418
421 425
428 429
435 437
443 444
445 446
450 454
458 465
466 467
470 472
475 477
478 480
485 486
487 494
496 496
506 516
517 520
521 525
527 528
533 534
537 538
540 542
543 545
546 546
550 550
552 553
555 557
561 565
566 567
569 572
573 573
SQ SEQUENCE 602 AA; 63953 MW; 982EF3245D87C43E CRC64;

Query Match 11.5%; Score 90; DB 1; Length 602;
Best Local Similarity 33.3%; Pred. No. 1.9; 16; Indels 34; Gaps 3;
Matches 28; Conservative 6; Mismatches

43 YDQLVTRVTHMAHALQ-----SDAR-----KSETTTTQSGYNG 78
254 YDQLVTRVTHMAHALGFGSPFFEDARIVANVPNVGRKNFDPVINSSTAVAKAREQVG 313
79 ADV-----GQGDADNSTIEL 92

```

DR MIM; 138248; --; C:plasma membrane; TAS.  
 DR GO; GO:0005886; F:glutamate receptor activity; TAS.  
 DR GO; GO:0008066; F:glutamate selective glutamate receptor activity; TAS.  
 DR GO; GO:0015277; F:kinase activity; TAS.  
 DR GO; GO:0007165; P:signal transduction; TAS.  
 DR GO; GO:0007268; P:synaptic transmission; TAS.  
 DR InterPro; IPR001828; ANF receptor.  
 DR InterPro; IPR001320; Ion\_glu\_receptor.  
 DR InterPro; IPR001622; K-channel\_pore.  
 DR InterPro; IPR001508; NMDA receptor.  
 DR InterPro; IPR001311; SBP/Glu receptor.  
 DR Pfam; PF01094; ANF receptor; 1.  
 DR Pfam; PF00060; lig\_chan; 1.  
 DR PRINTS; PR00177; NMDARECEPTOR.  
 DR SMART; SM00079; PBP; 1.  
 KW Receptor; Postsynaptic membrane; Ionic channel; Glycoprotein; Signal;  
 Transmembrane; Alternative splicing.  
 FT SIGNAL 1 18 POTENTIAL.  
 FT CHAIN 19 906 GLUTAMATE RECEPTOR 1.  
 FT DOMAIN 19 536 EXTRACELLULAR (POTENTIAL).  
 FT TRANSMEM 537 557 POTENTIAL.  
 FT TRANSMEM 585 605 POTENTIAL.  
 FT TRANSMEM 618 638 POTENTIAL.  
 FT TRANSMEM 806 826 POTENTIAL.  
 FT VARSPLIC 758 758  
 FT VARSPLIC 768 768  
 FT VARSPLIC 772 772  
 FT VARSPLIC 778 778  
 FT VARSPLIC 790 793  
 FT CONFLICT 345 345 R -> A (IN REF. 2).  
 FT CONFLICT 375 375 S -> G (IN REF. 2 AND 3).  
 FT CONFLICT 863 863 S -> SA (IN REF. 3).  
 FT CONFLICT 865 867 AGA -> TAP (IN REF. 3).  
 SQ SEQUENCE 906 AA; 101536 MW; 03EALB06D100BEEF CRC64;  
 Query Match 11.5%; Score 90; DB 1; Length 906;  
 Best Local Similarity 22.6%; Pred No; 3;  
 Matches 45; Conservative 26; Mismatches 56; Indels 72; Gaps 10;  
 QY 6 VAAFAIVVSGSALAGVFPWQGGNGHNGGSSGPDYDQL-----VTRVVTHE-----M 55  
 Db 250 VTGFLVNVYDTIPAKIMQW-----KNSDARDTRVDWPKPKYTSALTYDGVKVM 300  
 QY 56 AHALQSDARKSETITQSGYNGADV-----GOGAD-----NSTIELTO 94  
 Db 301 AEAFQSLRQ---RIDISRRGNAGCLANPVPWGGIDIQALQOVRPEGLTGNVQNE 357  
 QY 95 NGRPNRAT-----IDWNKNSDI---TVGQYGGNNAALVNQT-----ASD 132  
 Db 358 KGRRTVYTLHVEMKHSIRKIGYWNEDDKFPAATDAQAGDSSVQNRIVYVITILED 417  
 QY 133 SSVWVRQVFGNNATANQY 151  
 Db 418 PYVMLKK-----NANQF 429  
 RESULT 13  
 OMPB\_RICCN STANDARD; PRT; 1655 AA.  
 AC Q9KKA3; Q9KKA3; Q9KKA3; Q9KKA3; Q9KKA3; Q9KKA3; Q9KKA3; Q9KKA3; Q9KKA3; Q9KKA3;  
 DT 16-OCT-2001 (Rel. 40, Created)  
 DT 16-OCT-2001 (Rel. 40, Last sequence update)  
 DT 28-FEB-2003 (Rel. 41, Last annotation update)  
 DE Outer membrane protein B precursor (168 kDa surface-layer protein)  
 DE (Surface protein antigen) (Cell surface antigen 5) (Scas) (rOmpB)  
 DE (rOmpB) (Contains: 120 kDa surface-exposed protein (Surface protein  
 antigen) (120 kDa outer membrane protein ompB); 32 kDa beta peptide).  
 GN OMPB OR RC1085.

OS Rickettsia conorii.  
 OC Bacteria; Proteobacteria; Alphaproteobacteria; Rickettsiales;  
 OC Rickettsiaceae; Rickettsiae; Rickettsia.  
 RN NCBI\_TaxID=781;  
 (1)  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=Malish 7;  
 RX MEDLINE=21442074; PubMed=11557893;  
 RA Ogata H., Audic S., Renesto-Audiffren P., Fournier P.-E., Barbe V.,  
 RA Samson D., Roux V., Cossart P., Weissenbach J., Claverie J.-M.,  
 RA Raoult D.;  
 RT "Mechanisms of evolution in Rickettsia conorii and R. prowazekii";  
 RL Science 293:2093-2098(2001).  
 RN (2)  
 RP SEQUENCE OF 33-1649 FROM N.A.  
 RC STRAIN=Indian tick typhus, and Malish 7;  
 RX MEDLINE=20393643; PubMed=10939649;  
 RA Roux V., Raoult D.;  
 RT "Phylogenetic analysis of members of the genus Rickettsia using the  
 gene coding the outer-membrane protein ompB (ompB).";  
 RL Int. J. Syst. Evol. Microbiol. 50:1449-1455(2000).  
 RN (3)  
 RP SEQUENCE OF 353-1655 FROM N.A.  
 RC STRAIN=Malish 7;  
 RA Stenos J., Walker D.;  
 RT "The rickettsial outer membrane protein A and B genes of Rickettsia  
 australis, the most divergent rickettsia of the spotted fever group.";  
 RL Submitted (MAY-1999) to the EMBL/GenBank/DBJ databases.  
 CC -!- FUNCTION: THE 120 kDa SURFACE-EXPOSED PROTEIN IS A MAJOR  
 CC STRUCTURAL PROTEIN WHICH MAY PLAY A ROLE AS A RICKETTSIAL  
 CC VIRULENCE FACTOR AND/OR IMMUNOGEN DURING INFECTION (BY  
 CC similarity).  
 CC -!- FUNCTION: THE 32 kDa BETA PEPTIDE MAY SERVE AS A MEMBRANE ANCHOR  
 CC (By similarity).  
 CC -!- SUBCELLULAR LOCATION: Cell wall. This bacterium is covered by a S-  
 CC layer with hexagonal symmetry (By similarity).  
 CC -!- SIMILARITY: BELONGS TO THE RICKETTSIAE OMPA/OMP FAMILY.  
 -----  
 CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 -----  
 DR EMBL; AE008659; AAL03623.1; -;  
 DR EMBL; AF123721; AAF34124.1; -;  
 DR EMBL; AF123726; AAF34129.1; -;  
 DR EMBL; AF149110; AAD39533.1; -;  
 DR PIR; E97835; E97835.  
 DR InterPro; IPR006315; Autotransport.  
 DR InterPro; IPR005546; Autotransporter.  
 DR Pfam; PF03797; Autotransporter; 1.  
 DR TIGRFAMs; TIGR01414; autotrans\_barl; 2.  
 KW Antigen; S-layer; Cell wall; Complete proteome.  
 FT CHAIN 1 1334 120 kDa SURFACE-EXPOSED PROTEIN.  
 FT CHAIN 1335 1655 32 kDa BETA PEPTIDE.  
 FT VARIANT 61 61 P -> A (IN STRAIN INDIAN TICK TYPHUS).  
 FT VARIANT 75 75 G -> S (IN STRAIN INDIAN TICK TYPHUS).  
 FT VARIANT 78 78 K -> N (IN STRAIN INDIAN TICK TYPHUS).  
 FT VARIANT 251 251 V -> A (IN STRAIN INDIAN TICK TYPHUS).  
 FT VARIANT 413 413 N -> D (IN STRAIN INDIAN TICK TYPHUS).  
 FT VARIANT 959 959 I -> V (IN STRAIN INDIAN TICK TYPHUS).  
 FT VARIANT 988 988 A -> T (IN STRAIN INDIAN TICK TYPHUS).  
 FT VARIANT 1139 1139 R -> L (IN STRAIN INDIAN TICK TYPHUS).  
 FT VARIANT 353 354 KD -> CH (IN REF. 3).  
 FT CONFLICT 776 776 F -> S (IN REF. 3).  
 FT CONFLICT 1159 1159 E -> D (IN REF. 3).  
 FT CONFLICT 1177 1177 G -> S (IN REF. 3).  
 FT CONFLICT 1492 1492 H -> R (IN REF. 3).  
 SQ SEQUENCE 1655 AA; 168342 MW; E49E19377D5FCE37 CRC64;

Query Match 11.3%; Score 88.5; DB 1; Length 1655;  
Best Local Similarity 25.2%; Pred. No. 8.1;  
Matches 41; Conservative 17; Mismatches 60; Indels 45; Gaps 7;  
13 VVSGSALAGVVPQGGGNGHNGSGDPDYDQLVTRVVTTHMAHALQSDARKSETTITQ 72  
505 LVGGALAGTITLDSAITIGDIGNAGGAALQGIT-----LANDATK---TLTL 551  
73 SGY-----GNGADVGGGANSIELTQNGFRNATID-----QWNAKNS 111  
552 GGANIGAGGTINFGANGGTIKLTST--QNNIVVDFDLAATQGTGVVDASSITNAQTL 609  
112 DIT--VGQVGNNAAL-----VNQTASDSVMVRQVGFNN 145  
610 TINGKIGTVGANKTLQFNIGSSKTVLSDGVAINELVIGNN 652  
RESULT 14  
GP63 LEICH STANDARD; PRT; 599 AA.  
AC P15706;  
DT 01-APR-1990 (Rel. 14, Created)  
DT 01-APR-1990 (Rel. 14, Last sequence update)  
DT 10-OCT-2003 (Rel. 42, Last annotation update)  
DE Leishmanolysin precursor (EC 3.4.24.36) (Cell surface protease)  
DE (Major surface glycoprotein) (GP63 protein) (Promastigote surface endopeptidase).  
GN Leishmania chagasi.  
OS Eukaryota; Euzoenozoa; Kinetoplastida; Trypanosomatidae; Leishmania.  
OC NCBI\_TaxID=44271;  
OX [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=90205976; PubMed=2320059;  
RA Miller R.A., Reed S.G., Parsons M.;  
RT "Leishmania gp63 molecule implicated in cellular adhesion lacks an Arg-Gly-Asp sequence."  
RL Mol. Biochem. Parasitol. 39:267-274 (1990).  
RN [2]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=92112918; PubMed=1370484;  
RA Ramamorthy R., Donelson J.E., Paetz K.E., Maybodi M., Roberts S.C., Wilson M.E.;  
RT "Three distinct RNAs for the surface protease gp63 are differentially expressed during development of Leishmania donovani chagasi promastigotes to an infectious form."  
RL J. Biol. Chem. 267:1888-1895 (1992).  
CC -1- FUNCTION: Has an integral role during the infection of macrophages in the mammalian host.  
CC -1- CATALYTIC ACTIVITY: Preference for hydrophobic residues at P1 and P1', and basic residues at P2 and P3'. A model nonapeptide is cleaved at -Ala-Tyr-Lys-Lys-.  
CC -1- COFACTOR: Binds 1 zinc ion per subunit (By similarity).  
CC -1- SUBCELLULAR LOCATION: Attached to the membrane by a GPI-anchor.  
CC -1- SIMILARITY: Belongs to peptidase family M8.  
CC  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (see <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC  
CC EMBL; M80672; AAA29238.1; -;  
CC EMBL; M28527; AAA29235.1; -;  
CC PIR; A44951; A44951.  
CC HSP; P08148; 11ML.  
CC MEROPS; M08.001; -;  
CC InterPro; IPR006025; Rept\_M\_Zn\_BS.  
CC InterPro; IPR001577; Peptidase\_M8.  
CC Pfam; PF01457; Peptidase\_M8; 1.  
CC PRINTS; PR00782; LSHMANOLYSIN.

DR PROSITE; PS00142; ZINC\_PROTEASE; 1.  
KW Hydrolase; Metalloprotease; Glycoprotein; Metal-binding; Zinc;  
KW Zymogen; Signal; Cell adhesion; GPI-anchor; Lipoprotein.  
FT SIGNAL 1 39 POTENTIAL.  
FT PROPEP 40 97 ACTIVATION PEPTIDE.  
FT CHAIN 98 574 LEISHMANOLYSIN.  
FT PROPEP 575 599 REMOVED IN MATURE FORM (BY SIMILARITY).  
FT METAL 261 261 ZINC (CATALYTIC) (BY SIMILARITY).  
FT ACT SITE 262 262 BY SIMILARITY.  
FT METAL 265 265 ZINC (CATALYTIC) (BY SIMILARITY).  
FT METAL 331 331 ZINC (CATALYTIC) (BY SIMILARITY).  
FT DISULFID 122 139 BY SIMILARITY.  
FT DISULFID 188 227 BY SIMILARITY.  
FT DISULFID 311 383 BY SIMILARITY.  
FT DISULFID 390 452 BY SIMILARITY.  
FT DISULFID 403 422 BY SIMILARITY.  
FT DISULFID 412 486 BY SIMILARITY.  
FT DISULFID 463 507 BY SIMILARITY.  
FT DISULFID 512 562 BY SIMILARITY.  
FT DISULFID 532 555 BY SIMILARITY.  
FT CARBOHYD 297 297 N-LINKED (GLCNAC...) (POTENTIAL).  
FT CARBOHYD 394 394 N-LINKED (GLCNAC...) (POTENTIAL).  
FT LIPID 574 574 GPI-anchor amidated asparagine (By similarity).  
SQ SEQUENCE 599 AA; 63848 MW; 746730A8E2A2E7C CRC64;  
Query Match 11.3%; Score 88; DB 1; Length 599;  
Best Local Similarity 46.7%; Pred. No. 2.8;  
Matches 21; Conservative 3; Mismatches 11; Indels 10; Gaps 1;  
43 YDQLVTRVVTTHMAHALQSDARKSETTITQSGYNGADVGGQADN 87  
251 YDQLVTRVVTTHMAHALGFSV-----GPFEGARILEISIN 285  
RESULT 15  
OVO\_DROME STANDARD; PRT; 1028 AA.  
AC P51521; Q9XZU4;  
DT 01-OCT-1996 (Rel. 34, Created)  
DT 01-OCT-1996 (Rel. 34, Last sequence update)  
DT 10-OCT-2003 (Rel. 42, Last annotation update)  
DE Ovo protein (Shaven baby protein).  
GN OVO OR SVB.  
OS Drosophila melanogaster (Fruit fly).  
OC Eukaryota; Metazoa; Arthropoda; Insecta; Pterygota;  
OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;  
OC Ephydroidea; Drosophilidae; Drosophila.  
OX NCBI\_TaxID=7227;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC TISSUE-Ovary.  
RX MEDLINE=95021209; PubMed=7935398;  
RA Garfinkel M.D., Wang J., Liang Y., Mahowald A.P.;  
RT "Multiple products from the shavenbaby-ovo gene region of Drosophila melanogaster: relationship to genetic complexity."  
RT NCBI Cell. Biol. 14:6809-6818 (1994).  
RN [2]  
RP SEQUENCE FROM N.A.  
RC STRAIN-Oregon-R;  
RX MEDLINE=91293102; PubMed=1712294;  
RA Mevel-Ninio M.T.M., Terracol R., Kafatos F.C.;  
RT "The ovo gene of Drosophila encodes a zinc finger protein required for female germ line development."  
RL EMBO J. 10:2259-2266 (1991).  
CC -1- FUNCTION: REQUIRED FOR SURVIVAL AND DIFFERENTIATION OF FEMALE GERM LINE CELLS. PLAYS A ROLE IN GERM LINE SEX DETERMINATION.  
CC -1- SUBCELLULAR LOCATION: Nuclear (Potential).  
CC -1- DEVELOPMENTAL STAGE: FIRST APPEARS IN THE GERMARINUM AND ACCUMULATES IN NURSE CELLS DURING OOGENESIS. STORED IN THE EGG, BUT IS RAPIDLY LOST IN THE EMBRYOS EXCEPT FOR ITS CONTINUED PRESENCE IN THE GERM LINE PRECURSOR POLE CELLS.  
CC -1- SIMILARITY: Contains 4 C2H2-type zinc fingers.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----

DR EMBL; U11383; AAB60216.1; -;  
DR EMBL; X59772; CAB36921.1; ALT\_SEQ.  
DR PIR; A56038; A56038.  
DR HSP; P07248; 2ADR.  
DR TRANSFAC; T00669; -;  
DR FlyBase; FBgn003028; ovo.  
DR InterPro; IPR007087; Znf\_C2H2.  
DR Pfam; PF00096; zf-C2H2; 3.  
DR SMART; SM00355; Znf\_C2H2; 4.  
DR PROSITE; PS00028; ZINC\_FINGER\_C2H2\_1; 3.  
DR PROSITE; PS0157; ZINC\_FINGER\_C2H2\_2; 3.  
KW Zinc-finger; Metal-binding; DNA-binding; Repeat; Nuclear protein;  
KW Transcription regulation.  
FT DOMAIN 62 66 POLY-ALA.  
FT DOMAIN 72 77 POLY-GLY.  
FT DOMAIN 80 85 POLY-GLY.  
FT DOMAIN 98 108 POLY-GLY.  
FT DOMAIN 144 152 POLY-HIS.  
FT DOMAIN 153 159 POLY-ASN.  
FT DOMAIN 336 339 POLY-GLN.  
FT DOMAIN 347 353 POLY-GLN.  
FT DOMAIN 357 361 POLY-GLN.  
FT DOMAIN 410 414 POLY-GLN.  
FT DOMAIN 418 422 POLY-GLN.  
FT DOMAIN 426 432 POLY-GLN.  
FT DOMAIN 445 453 POLY-GLN.  
FT DOMAIN 456 459 POLY-GLN.  
FT DOMAIN 466 474 POLY-GLN.  
FT DOMAIN 497 517 POLY-ALA.  
FT DOMAIN 524 529 POLY-SER.  
FT DOMAIN 549 558 POLY-ALA.  
FT DOMAIN 639 651 POLY-ALA.  
FT DOMAIN 717 725 POLY-ALA.  
FT DOMAIN 797 802 POLY-GLN.  
FT DOMAIN 820 823 POLY-GLN.  
FT DOMAIN 826 832 POLY-GLN.  
FT ZN\_FING 874 896 C2H2-TYPE 1.  
FT ZN\_FING 902 924 C2H2-TYPE 2.  
FT ZN\_FING 930 953 C2H2-TYPE 3.  
FT ZN\_FING 969 992 C2H2-TYPE 4.  
FT CONFLICT 647 647 A -> R (IN REF. 2).  
SQ SEQUENCE 1028 AA; 110620 MW; D7068BB2C0F677 CRC64;

Query Match 11.2%; Score 87.5; DB 1; Length 1028;  
Best Local Similarity 25.6%; Pred. No. 5.7;  
Matches 40; Conservative 11; Mismatches 58; Indels 47; Gaps 6;  
QY 3 LKVAFAAIVVSGSALAGVFWGGGNGNHCNCGSGDPDYLTVRVVTHMAHALQSD 62  
DB 59 LQNAAAAAYIMSAGSG-----GGGCTGNGGGGASGP----- 89  
QY 63 ARKSETTITQSGYNGADYVGQADN-----STIELTQNGFERNATIDQWNAKNSDI--- 113  
DB 90 ---GGGPSANSGGGGGGGGNGYVCGGVGPNNSLDGNLLNFASVSNYNESKSFNH 146  
QY 114 -TVQYGGNNAALVQNTASDSVMVQVGFGNATA 148  
DB 147 HHHHQHNNN-----NNGGQTSWVGHPF--YGCNPSA 177

Search completed: August 2, 2004, 14:49:29  
Job time : 5.3 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

# OM protein - protein search, using sw model

Run on: August 2, 2004, 14:39:17 ; Search time 29.7 Seconds  
(without alignments)  
1604.150 Million cell updates/sec

Title: US-09-543-407-18

Perfect score: 780

Sequence: 1 MKLLKVAFAAIVSGSALA.....DSSVMVRQVGFNNATANQY 151

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1017041 seqs, 315518202 residues

Total number of hits satisfying chosen parameters: 1017041

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

SPTREMBL 25:\*

- 1: sp\_archaea:\*
- 2: sp\_bacteria:\*
- 3: sp\_fungi:\*
- 4: sp\_human:\*
- 5: sp\_invertebrate:\*
- 6: sp\_mammal:\*
- 7: sp\_mhc:\*
- 8: sp\_organelle:\*
- 9: sp\_phage:\*
- 10: sp\_plant:\*
- 11: sp\_rodent:\*
- 12: sp\_virus:\*
- 13: sp\_vertebrate:\*
- 14: sp\_unclassified:\*
- 15: sp\_virus:\*
- 16: sp\_bacteriaph:\*
- 17: sp\_archaeosp:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	682	87.4	152	2	O33802 salmonella
2	593.5	76.1	150	2	Q7X243 citrobacter
3	550	70.5	149	2	Q7X240 citrobacter
4	498.5	63.9	152	16	Q8CW63 escherichia
5	428.5	54.9	150	2	Q7X237 enterobacte
6	303	38.8	76	2	Q54069 salmonella
7	122	15.6	29	2	Q9S3J5 escherichia
8	114	14.6	1748	5	Q94821 tetrahymena
9	113	14.5	502	16	Q8EIH4 shewanella
10	110.5	14.2	151	2	Q7X244 citrobacter
11	108	13.8	171	16	Q89J13 bradyrhizob
12	107.5	13.8	151	16	Q7UC21 shigella fl
13	107.5	13.8	160	16	Q8CW64 escherichia
14	107.5	13.8	160	16	Q83RU7 shigella fl
15	106.5	13.7	1209	16	Q89CK5 bradyrhizob
16	105.5	13.5	91	2	Q9S3J8 escherichia

17	101.5	13.0	154	16	Q89J15
18	101	12.9	262	5	Q9VIX5
19	100	12.8	179	2	O33801 salmonella
20	99	12.7	130	16	Q89J14
21	98	12.6	157	16	Q88HG0
22	98	12.6	362	16	Q8EV84
23	98	12.6	3659	16	Q98LN6
24	97.5	12.5	151	2	Q7X238
25	97.5	12.5	152	2	Q7X241
26	97	12.4	368	16	Q8EWD6
27	97	12.4	1422	16	Q8EFU3
28	97	12.4	2035	2	Q9XCJ4
29	97	12.4	2039	16	Q8ZNS7
30	96.5	12.4	1765	16	Q7V8S5
31	96	12.3	490	16	Q8EYI9
32	95.5	12.2	145	16	Q8UGN9
33	95.5	12.2	1613	2	Q9KXB2
34	95.5	12.2	3501	16	Q8Y106
35	95.5	12.2	3552	16	Q8XSD6
36	94.5	12.1	153	16	Q89J16
37	94.5	12.1	348	13	O93397
38	94	12.1	480	16	Q89EV2
39	93.5	12.0	139	16	Q8EIH3
40	93.5	12.0	287	5	Q9VIX6
41	93.5	12.0	582	16	P71868
42	93.5	12.0	582	16	Q7TW98
43	93.5	12.0	1615	2	Q9KKA8
44	93	11.9	141	16	Q8U6P1
45	92	11.8	353	16	Q8EV92

## ALIGNMENTS

### RESULT 1

O33802 PRELIMINARY; PRT; 152 AA.

AC O33802;  
 DT 01-JAN-1998 (TRENBLrel. 05, Created)  
 DT 01-JAN-1998 (TRENBLrel. 05, Last sequence update)  
 DT 01-DEC-2001 (TRENBLrel. 19, Last annotation update)  
 DE AGFA protein (Fragment).  
 GN AGFA.  
 OS Salmonella typhimurium.  
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;  
 OC Enterobacteriaceae; Salmonella.  
 OX NCBI\_TaxID=602;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=98053981; PubMed=9393832;  
 RA Sukupolvi S.S., Lorentz R.G., Gordon J.I., Bian Z., Pfeifer J.D.,  
 RA Normark S.J., Rhen M.;  
 RT "Expression of thin, aggregative fimbriae promotes interaction of  
 RT Salmonella typhimurium SR-11 with mouse small intestinal epithelial  
 RT cells.";  
 RL Infect. Immun. 65:5320-5325(1997).  
 DR EMBL; AJ000514; CAA04151.1; -;  
 FT NON TER 152 152  
 SQ SEQUENCE 152 AA; 15401 MW; 9DA7DADC2364B006 CRC64;

Query Match 87.4%; Score 682; DB 2; Length 152;  
 Best Local Similarity 89.4%; Pred. No. 1.7e-46;  
 Matches 135; Conservative 4; Mismatches 13; Indels 0; Gaps 0;  
 QY 1 MKLLKVAFAAIVSGSALAAGVVPQWGGGNGHNGSGSPDYLQVTRVTHEVAHALQ 60  
 DB 1 MKLLKVAFAAIVSGSALAAGVVPQWGGGNGHNGSGSPDSTLSIYQGSANAALALQ 60  
 QY 61 SPARKSETTITQSGVNGADVCGQADNSTIETLQNGFNATTIDQWNAKNSDITVQYGG 120  
 DB 61 SPARKSETTITQSGVNGADVCGQADNSTIETLQNGFNATTIDQWNAKNSDITVQYGG 120  
 QY 121 NNAALVNQTSBSSVMVRQVGFNNATANQY 151

```

Db      121 NNAALVNQTASDSSVMVRQVGFNNAPANQY 151
|||||
RESULT 2
Q7X243 PRELIMINARY; PRT; 150 AA.
AC Q7X243;
DT 01-OCT-2003 (TrEMBLrel. 25, Created)
DT 01-OCT-2003 (TrEMBLrel. 25, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Curlin-csgA protein.
GN CSGA.
OS Citrobacter sp. Fec2.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Citrobacter.
OX NCBI_TaxID=213763;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=fec2;
RA Zogaj X., Bokranz W., Nimitz M., Romling U.;
RT "Production of Cellulose and Curli Fimbriae by Members of the Family
RT Enterobacteriaceae Isolated from the Human Gastrointestinal Tract.";
RL Infect. Immun. 72:4151-4158(2003).
DR EMBL; AJ515700; CAD56672.1; -
SQ SEQUENCE 150 AA; 15016 MW; 1D7141B8D6973DC6 CRC64;

Query Match 76.1%; Score 593.5; DB 2; Length 150;
Best Local Similarity 78.1%; Pred. No. 1.7e-39;
Matches 118; Conservative 14; Mismatches 18; Indels 1; Gaps 1;

QY 1 MLLKVAFAAIVVSGSALAGVVPQW-GGNGHGGSGNPGDSSLSIYQYGSNNNALQ 60
|||||
Db 1 MLLQVAFAAIVVSGSALAGVVPQWGGGSGNPGDSSLSIYQYGSNNNALQ 59
|||||
QY 61 SDARKSETTITQSGYNGADVGQADNSTIETQNGFRNNATIDOWNAKNSDITVGYGG 120
|||||
Db 60 SDARKSDTTHQNGFGNGADVGQSDNSTIDLTQNGFKNNATIDQWNGKNSDITVGYGG 119
|||||

QY 121 NNAALVNQTASDSSVMVRQVGFNNATANQY 151
|||||
Db 120 HNAALVNQTASDSSVLVHQVGFNNATANQY 150
|||||

RESULT 3
Q7X240 PRELIMINARY; PRT; 149 AA.
AC Q7X240;
DT 01-OCT-2003 (TrEMBLrel. 25, Created)
DT 01-OCT-2003 (TrEMBLrel. 25, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Curlin-csgA protein.
GN CSGA.
OS Citrobacter freundii.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Citrobacter.
OX NCBI_TaxID=546;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=fec4;
RA Zogaj X., Bokranz W., Nimitz M., Romling U.;
RT "Production of Cellulose and Curli Fimbriae by Members of the Family
RT Enterobacteriaceae Isolated from the Human Gastrointestinal Tract.";
RL Infect. Immun. 72:4151-4158(2003).
DR EMBL; AJ515701; CAD56675.1; -
SQ SEQUENCE 149 AA; 15260 MW; 946DD52017F648FD CRC64;

Query Match 70.5%; Score 550; DB 2; Length 149;
Best Local Similarity 72.2%; Pred. No. 4.5e-36;
Matches 109; Conservative 20; Mismatches 20; Indels 2; Gaps 1;

QY 1 MLLKVAFAAIVVSGSALAGVVPQWGGGNGHGGSGNPGDSSLSIYQYGSNNNALQ 60
|||||

```





```
Db 1681 -SSGNGQ----TGGHGSN---DNQOQOQNTGGGWSGSSNS---NQTWNSSS-----1722
Qy 142 FGNATYAN 149
Db 1723 WGSNNQAS 1730

RESULT 9
Q8EIH4 PRELIMINARY; PRT; 502 AA.
AC Q8EIH4
DT 01-MAR-2003 (TremBrel. 23, Created)
DT 01-MAR-2003 (TremBrel. 23, Last sequence update)
DT 01-MAR-2003 (TremBrel. 23, Last annotation update)
DE Conserved hypothetical protein.
GN S00865.
OS Shewanella oneidensis.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Alteromonadales;
OC Alteromonadaceae; Shewanella.
OX NCBI_TaxID=70863;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=MR-1;
RX MEDLINE=2297686; PubMed=12368813;
RA Heidelberg J.F., Paulsen I.T., Nelson K.E., Gaidos E.J., Nelson W.C.,
RA Read T.D., Eisen J.A., Seshadri R., Ward N., Methe B., Clayton R.A.,
RA Meyer R.T., Tsapin A., Scott J., Beanan M., Brinkac L., Daugherty S.,
RA DeBoy R.T., Dodson R.J., Durkin A.S., Haft D.H., Kolonay J.F.,
RA Madupu R., Peterson J.D., Unayam L.A., White O., Wolf A.M.,
RA Vamathevan J., Weidman J., Impraim M., Lee K., Berry K., Lee C.,
RA Mueller J., Khouri H., Gill J., Utterback T.R., McDonald L.A.,
RA Feldblyum T.V., Smith H.O., Venter J.C., Nealon K.H., Fraser C.M.;
RT "Genome sequence of the dissimilatory metal ion-reducing bacterium
RT Shewanella oneidensis."
RL Nat. Biotechnol. 20:1118-1123(2002).
DR EMBL; AB015532; AAN53941.1; -
DR TIGR; S00865; -
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 502 AA; 52441 MW; D08CA23D6C46B62D CRC64;

Query Match 14.5%; Score 113; DB 16; Length 502;
Best Local Similarity 26.0%; Pred. No. 0.64;
Matches 40; Conservative 20; Mismatches 58; Indels 36; Gaps 6;

Qy 29 GGNHNG-----GGN-----SSGPDYDQVTR-----VVTHEMAHAL 59
Db 231 GDNHTGFVYALAGSENDISMEQEGSNNTAYLSNTTGDNTVDITQDGSNTVGDLSIADI 290

Qy 60 QSDARKSETTITQSGYNGADVQCGADNSTIELTQNGFNNTATIDQWNAKNSDITVQYQ 119
Db 291 QGD--DNDITIKQKGSNGAEFGWGSNDVDLKQGDGFANFATFGAYGTDN--DFDLSSKG 347

Qy 120 GNNALVNQTSASSVMVQVQFGN-----NATAN 149
Db 348 DNNELVAFATGEDNSIEIQEGDANFAYDATGN 381

RESULT 10
Q7X244 PRELIMINARY; PRT; 151 AA.
AC Q7X244
DT 01-OCT-2003 (TremBrel. 25, Created)
DT 01-OCT-2003 (TremBrel. 25, Last sequence update)
DT 01-OCT-2003 (TremBrel. 25, Last annotation update)
DE Nucleation component of curlin monomers.
GN CSGB.
OS Citrobacter sp. Fec2.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Citrobacter.
OX NCBI_TaxID=213763;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Fec2;

RA Zogaj X., Bokranz W., Nimtz M., Romling U.;
RT "Production of Cellulose and Curli Fimbriae by Members of the Family
RT Enterobacteriaceae Isolated from the Human Gastrointestinal Tract.";
RL Infect. Immun. 72:4151-4158(2003).
DR EMBL; AJ515700; CAD56671.1; -.
SQ SEQUENCE 151 AA; 16158 MW; BD00AF57E1400704 CRC64;

Query Match 14.2%; Score 110.5; DB 2; Length 151;
Best Local Similarity 25.0%; Pred. No. 0.24;
Matches 29; Conservative 26; Mismatches 46; Indels 15; Gaps 3;

Qy 47 VTRVVTHEVAHALQSDARKSETTITQSGYNGADVQCGADNSTIELTQNGFN-----98
Db 18 IASATSYDLAH---SEYNFAVNELSKSFQAIIIGVGTNNNAKRWQEGSKLLSVSQE 74

Qy 99 ---NNATIDQWNAKNSDITVQYQGNNAALVNQTSASSVMVQVQFGNNTATQY 151
Db 75 GGSNRAKVDQSGAYNF-AYIAQSGHSDASISQSNYGNNTAMIIQKSGNKANITQY 129

RESULT 11
Q89J13 PRELIMINARY; PRT; 171 AA.
AC Q89J13
DT 01-JUN-2003 (TremBrel. 24, Created)
DT 01-JUN-2003 (TremBrel. 24, Last sequence update)
DT 01-JUN-2003 (TremBrel. 24, Last annotation update)
DE CSGA protein.
GN CSGA OR BL5300.
OS Bradyrhizobium japonicum.
OC Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;
OC Bradyrhizobiaceae; Bradyrhizobium.
OX NCBI_TaxID=375;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=USDA 110;
RX MEDLINE=2248498; PubMed=12597275;
RA Kaneko T., Nakamura Y., Sato S., Minamisawa K., Uchiumi T.,
RA Sasamoto S., Watanabe A., Idegawa K., Iriguchi M., Kawashima K.,
RA Kohara M., Matsumoto M., Shimpō S., Tsuruoka H., Wada T., Yamada M.,
RA Tabata S.;
RT "Complete genomic sequence of nitrogen-fixing symbiotic bacterium
RT Bradyrhizobium japonicum USDA110.";
RL DNA Res. 9:189-197(2002).
DR EMBL; AP005954; BAC50565.1; -.
KW Complete proteome.
SQ SEQUENCE 171 AA; 17448 MW; 995DB08C01498381 CRC64;

Query Match 13.8%; Score 108; DB 16; Length 171;
Best Local Similarity 36.6%; Pred. No. 0.44;
Matches 34; Conservative 10; Mismatches 39; Indels 10; Gaps 3;

Qy 58 ALOSADARKSETTITQSGYNGADVQCGADNSTIELTQNGFNNTATIDQWNAKNSDITV 115
Db 52 ALSAQAQANTSTTVQVGLVNGSVTQNGLTNDSSSTTQIGILNGASTMQGTSSPS----107

Qy 116 GOYGGNNAALVNQTSASSVMVQVQFGNNTA 148
Db 108 ---LNNVTNVQAGVQNSATTCQVAFGNNGSA 136

RESULT 12
Q7UCZ1 PRELIMINARY; PRT; 151 AA.
AC Q7UCZ1
DT 01-OCT-2003 (TremBrel. 25, Created)
DT 01-OCT-2003 (TremBrel. 25, Last sequence update)
DT 01-OCT-2003 (TremBrel. 25, Last annotation update)
DE Minor curlin subunit.
GN CSGB OR S1108.
OS Shigella flexneri.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Shigella.
```

```

OX NCBI_TaxID=623;
RN SEQUENCE FROM N.A.
RP STRAIN=2457T / ATCC 700330 / Serotype 2a;
RX MEDLINE=22590274; PubMed=12704152;
RA Wei J., Goldberg M.B., Burland V., Venkatesan M.M., Deng W.,
RA Fournier G., Mayhew G.F., Plunkett G. III, Rose D.J., Darling A.,
RA Mau B., Perna N.T., Payne S.M., Runyan-Janecky L.J., Zhou S.,
RA Schwartz D.C., Blattner F.R.;
RT "Complete genome sequence and comparative genomics of Shigella
RL flexneri serotype 2a strain 2457T.";
DR EMBL; AB016981; AAP16542.1; -
SQ SEQUENCE 151 AA; 15868 MW; 5D5D266B964014A0 CRC64;

Query Match 13.8%; Score 107.5; DB 16; Length 151;
Best Local Similarity 35.4%; Pred. No. 0.42; Indels 1; Gaps 1;
Matches 29; Conservative 9; Mismatches 43;

QY 70 ITQSGYNGADVGQGDNSTIELTQNGFRNNATIDQNAKNSDITVGYGNNALVNQT 129
Db 49 IGQAGTNSAQLRQGGSKLLAVVAQEGSSNRRAKIDQTGYNL-AYIDQAGSANDASISQG 107

QY 130 ASDSSVMVRQVGFNNATANQY 151
Db 108 AYGNTAMIIQKSGNKANITQY 129

RESULT 13
Q8CW64 PRELIMINARY; PRT; 160 AA.
ID Q8CW64
AC Q8CW64;
DT 01-MAR-2003 (TReMBLrel. 23, Created)
DT 01-MAR-2003 (TReMBLrel. 23, Last sequence update)
DT 01-MAR-2003 (TReMBLrel. 23, Last annotation update)
DE Minor curlin subunit precursor.
GN CsgB OR C1305.
OS Escherichia coli 06.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
CC Enterobacteriaceae; Escherichia.
OX NCBI_TaxID=217992;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=O6:HI / CFT073 / ATCC 700928;
RX MEDLINE=22388234; PubMed=124711157;
RA Welch R.A., Burland V., Plunkett G. III, Redford P., Roesch P.,
RA Rasko D., Buckles E.L., Liou S.-R., Boutin A., Hackett J., Stroud D.,
RA Mayhew G.F., Rose D.J., Zhou S., Schwartz D.C., Perna N.T.,
RA Mobley H.L.T., Donnenberg M.S., Blattner F.R.;
RT "Extensive mosaic structure revealed by the complete genome sequence
RT of uropathogenic Escherichia coli.";
RL Proc. Natl. Acad. Sci. U.S.A. 99:17020-17024(2002).
DR EMBL; AE016759; AAN7978.1; -
KW Complete proteome.
SQ SEQUENCE 160 AA; 16963 MW; 49F68448D979B986 CRC64;

Query Match 13.8%; Score 107.5; DB 16; Length 160;
Best Local Similarity 35.4%; Pred. No. 0.45; Indels 1; Gaps 1;
Matches 29; Conservative 9; Mismatches 43;

QY 70 ITQSGYNGADVGQGDNSTIELTQNGFRNNATIDQNAKNSDITVGYGNNALVNQT 129
Db 58 IGQAGTNSAQLRQGGSKLLAVVAQEGSSNRRAKIDQTGYNL-AYIDQAGSANDASISQG 116

QY 130 ASDSSVMVRQVGFNNATANQY 151
Db 117 AYGNTAMIIQKSGNKANITQY 138

RESULT 14
Q83RU7 PRELIMINARY; PRT; 160 AA.
ID Q83RU7
AC Q83RU7;

```

```

DT 01-JUN-2003 (TReMBLrel. 24, Created)
DT 01-JUN-2003 (TReMBLrel. 24, Last sequence update)
DT 01-JUN-2003 (TReMBLrel. 24, Last annotation update)
DE Minor curlin subunit precursor, similar to CsgA.
GN CsgB OR SF1035.
OS Shigella flexneri.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
CC Enterobacteriaceae; Shigella.
OX NCBI_TaxID=623;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=301 / Serotype 2a;
RX MEDLINE=22272406; PubMed=12384590;
RA Jin Q., Yuan Z., Xu J., Wang Y., Shen Y., Lu W., Wang J., Liu H.,
RA Yang J., Yang F., Zhang X., Zhang J., Yang G., Wu H., Qu D., Dong J.,
RA Sun L., Xue Y., Zhao A., Gao Y., Zhu J., Kan B., Ding K., Chen S.,
RA Cheng H., Yao Z., He B., Chen R., Ma D., Qiang B., Wen Y., Hou Y.,
RA Yu J.;
RT "Genome sequence of Shigella flexneri 2a: insights into pathogenicity
RT through comparison with genomes of Escherichia coli K12 and O157.";
RL Nucleic Acids Res. 30:4432-4441(2002).
DR EMBL; AE015131; AAM42658.1; -
KW Complete proteome.
SQ SEQUENCE 160 AA; 16919 MW; 50269F5268D2A32F CRC64;

Query Match 13.8%; Score 107.5; DB 16; Length 160;
Best Local Similarity 35.4%; Pred. No. 0.45; Indels 1; Gaps 1;
Matches 29; Conservative 9; Mismatches 43;

QY 70 ITQSGYNGADVGQGDNSTIELTQNGFRNNATIDQNAKNSDITVGYGNNALVNQT 129
Db 58 IGQAGTNSAQLRQGGSKLLAVVAQEGSSNRRAKIDQTGYNL-AYIDQAGSANDASISQG 116

QY 130 ASDSSVMVRQVGFNNATANQY 151
Db 117 AYGNTAMIIQKSGNKANITQY 138

RESULT 15
Q89CK5 PRELIMINARY; PRT; 1209 AA.
ID Q89CK5
AC Q89CK5;
DT 01-JUN-2003 (TReMBLrel. 24, Created)
DT 01-JUN-2003 (TReMBLrel. 24, Last sequence update)
DT 01-JUN-2003 (TReMBLrel. 24, Last annotation update)
DE Bll7792 protein.
GN Bll7792.
OS Bradyrhizobium japonicum.
OC Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;
CC Bradyrhizobiaceae; Bradyrhizobium.
OX NCBI_TaxID=375;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=USDA 110;
RX MEDLINE=22484998; PubMed=12597275;
RA Kaneko T., Nakamura Y., Sato S., Minamisawa K., Uchiumi T.,
RA Sasamoto S., Watanabe A., Idesawa K., Iriguchi M., Kawashima K.,
RA Kohara M., Matsumoto M., Shimpo S., Tsuruoka H., Wada T., Yamada M.,
RA Tabata S.;
RT "Complete genomic sequence of nitrogen-fixing symbiotic bacterium
RT Bradyrhizobium japonicum USDA110.";
RL DNA Res. 9:189-197(2002).
DR EMBL; AF005963; BAC53057.1; -
KW Complete proteome.
SQ SEQUENCE 1209 AA; 118292 MW; 50F97581D524EE71 CRC64;

Query Match 13.7%; Score 106.5; DB 16; Length 1209;
Best Local Similarity 25.0%; Pred. No. 5.9; Indels 55; Gaps 8;
Matches 47; Conservative 20; Mismatches 66;

QY 9 FFAIVVSGSALAGVVPQWGGGNGHNGGNSGPDYDQLVTRVVTHEMAF----ALQSDAR 64
Db 934 YGNVTVTGSTGATTIAL---GNGNDTIDASG--YGNVITLNGNDIVHPGDSQTTAG 987

```

```
QY      65 KSETTITQSGYNGADVGGAD-----NSTIETQNGFRNNATIDQWN- 107
Db      988 NGNDLVTLSCYGNVLMGNGNDVWAGRDGANSVTLGDNNTVNL--GGMGNQITVSGTN 1045
QY      108 ----AKNSDITVCGYG-----GNNALVNO-----TASDSSVMVROVG 141
Db      1046 AIIAGSGSDNVWAGAGHDTIMLGGAANHVVLAGSQANVTNQIGQDVVTVNGGSDQFNFBG 1105
QY      142 FGNNATAN 149
Db      1106 FGQRAIN 1113
```

Search completed: August 2, 2004, 14:54:38  
Job time : 29.7 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: August 2, 2004, 14:35:42 ; Search time 44.9 Seconds  
(without alignments)  
950.215 Million cell updates/sec

Title: US-09-543-407-20

Perfect score: 774

Sequence: 1 MKLLKVAFAAIVVSGSALA.....DSSVMVRQVGFNNATANYQ 151

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1586107 seqs, 282547505 residues

Total number of hits satisfying chosen parameters: 1586107

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-Processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : A\_Geneseq\_29Jan04.\*

- 1: Geneseq1980s.\*
- 2: Geneseq1990s.\*
- 3: Geneseq2000s.\*
- 4: Geneseq2001s.\*
- 5: Geneseq2002s.\*
- 6: Geneseq2003as.\*
- 7: Geneseq2003bs.\*
- 8: Geneseq2004s.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	774	100.0	151	3	AAB36350
2	709	91.6	151	3	AAB36353
3	689	89.0	151	2	AAB74625
4	689	89.0	151	3	AAB36341
5	684	88.4	151	2	AAB36350
6	662	85.5	151	3	AAB36354
7	622	80.4	151	3	AAB36349
8	619	80.0	151	3	AAB36351
9	611	78.9	151	3	AAB36346
10	609	78.7	151	3	AAB36347
11	606	78.3	151	3	AAB36352
12	597	77.1	151	3	AAB36355
13	574	74.2	151	3	AAB36348
14	528	68.2	151	3	AAB36343
15	523	67.6	151	7	ABR82651
16	504	65.1	120	2	AAR2761
17	504	65.1	120	2	AAR23569
18	450	58.1	142	2	AAR2684
19	378	48.8	122	2	AAR2663
20	153	19.8	45	3	AAB36316
21	132	17.1	22	3	AAB36318
22	123	15.9	23	3	AAB36321
23	123	15.9	23	3	AAB36326
24	123	15.9	23	3	AAB36338
25	113	14.6	24	7	ABR82644

26	111	14.3	22	3	AAB36322	AAB36322 Salmonell
27	111	14.3	22	3	AAB36327	AAB36327 Salmonell
28	111	14.3	22	3	AAB36337	AAB36337 Salmonell
29	109	14.1	23	3	AAB36340	AAB36340 Salmonell
30	109	14.1	23	3	AAB36324	AAB36324 Salmonell
31	109	14.1	23	3	AAB36319	AAB36319 Salmonell
32	102	13.2	26	7	ABR82649	ABR82649 E. coli V
33	96.5	12.5	151	3	AAB36344	AAB36344 Escherich
34	96	12.4	19	3	AAB36323	AAB36323 Salmonell
35	96	12.4	19	3	AAB36336	AAB36336 Salmonell
36	96	12.4	19	3	AAB36328	AAB36328 Salmonell
37	95	12.3	24	7	ABR82647	ABR82647 E. coli C
38	92	11.9	23	3	AAB36331	AAB36331 Escherich
39	91	11.8	186	6	ABU21488	ABU21488 Protein e
40	91	11.8	502	2	AAB23212	AAB23212 Leishmani
41	90.5	11.7	677	4	ABG04318	ABG04318 Novel hum
42	90	11.6	24	7	ABR82642	ABR82642 E. coli N
43	89.5	11.6	423	4	ABG07164	ABG07164 Novel hum
44	89.5	11.6	447	3	AAG29728	AAG29728 Arabidops
45	89.5	11.6	468	3	AAG29727	AAG29727 Arabidops

#### ALIGNMENTS

#### RESULT 1

AAB36350  
ID AAB36350 standard; protein; 151 AA.

XX AAB36350;

DT 26-FEB-2001 (first entry)

DE Agfa::PT3#5 amino acid sequence SEQ ID NO:20.

XX Salmonella; agfa; chromosomal gene replacement; fimbrin; epitope;

KW vaccine; immune response; immunogen.

XX Salmonella enteritidis.

OS Escherichia coli.

OS Synthetic.

XX WO2000060102-A2.

PD 12-OCT-2000.

XX 05-APR-2000; 2000WO-CA000356.

XX 05-APR-1999; 99US-0127888P.

XX (UYVI-) UNIV VICTORIA.

PI White AP, Doran JL, Collison SK, Kay WW;

XX WPI; 2000-672631/65.

XX N-PSDB; AAC64626.

Recombinant agfa gene having a segment replaced by a foreign DNA sequence which encodes foreign epitope or antigen, expresses recombinant Agfa protein useful for eliciting immune response in animal.

PS Disclosure; Page 137; 139pp; English.

XX The present invention describes a recombinant agfa gene (I) where a segment of the gene has been replaced by a segment of a foreign DNA sequence which encodes a foreign epitope or antigen. Also described are: (1) use of thin aggregative fimbriae (SEF17/TAF) nucleation depended assembly system of strains of Salmonella, Escherichia coli and Enterobacteriaceae for the production of fimbriae comprising recombinant Agfa, CsgA and Agfa-homologue fimbrial subunits, respectively; (2) directing recombination of a recombinant gene into the chromosome of the homologous species; (3) directing recombination of a recombinant gene back into the chromosome of the homologous species, replacing the native

CC copy of that gene; and (4) eliciting an immune response in an animal,  
 CC comprising separating an amino acid polymer comprising a recombinant Agfa  
 CC protein containing a replacement segment or segments of foreign amino  
 CC acid sequence or sequences grown on a Salmonella, E. coli or  
 CC Enterobacteriaceae host cell, from the host cell and introducing the  
 CC polymer into the animal in conjunction with a carrier or diluent. (1) is  
 CC useful for the expression of recombinant Agfa protein which is useful for  
 CC eliciting an immune response in an animal. In a fibrial presentation  
 CC system the heterologous antigens are presented in high numbers (up to  
 CC 500,000 copies/cell), the hybrid fimbriae protein possesses both the  
 CC immunogenicity and adhesion properties relevant for an efficient live  
 CC vaccine, the carrier fimbrial subunit proteins are usually strong  
 CC immunogens, which may be important for directing an immune response  
 CC against the inserted epitope, and hybrid fimbriae are easy and  
 CC inexpensive to purify in large amount. The present sequence is given in  
 CC the exemplification of the present invention  
 XX  
 XX Sequence 151 AA;

Query Match 100.0%; Score 774; DB 3; Length 151;  
 Best Local Similarity 100.0%; Pred. No. 8.8e-71;  
 Matches 151; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 MKLLKVAAPAAIVVSGSALAGVVPQWGGGNGHNGGNSGPDSTLSIYQYGSANAALALQ 60  
 DB 1 MKLLKVAAPAAIVVSGSALAGVVPQWGGGNGHNGGNSGPDSTLSIYQYGSANAALALQ 60  
 QY 61 SDARKYDQLVTRVVTTHMAHAGGADNSTIETQNGFRNNATIDQNAKNSDITVQYGG 120  
 DB 61 SDARKYDQLVTRVVTTHMAHAGGADNSTIETQNGFRNNATIDQNAKNSDITVQYGG 120  
 QY 121 NNAALVNQTSADSSVMVRQVGFNNATANQY 151  
 DB 121 NNAALVNQTSADSSVMVRQVGFNNATANQY 151

RESULT 2  
 AAB36353  
 ID AAB36353 standard; protein; 151 AA.  
 AC AAB36353;  
 XX  
 XX 26-FEB-2001 (first entry)  
 XX Agfa::PT3#8 amino acid sequence SEQ ID NO:26.  
 XX Salmonella; agfa; chromosomal gene replacement; fimbria; epitope;  
 KW vaccine; immune response; immunogen.  
 XX Salmonella enteritidis.  
 OS Escherichia coli.  
 OS Synthetic.  
 XX WO200006102-A2.  
 XX 12-OCT-2000.  
 XX 05-APR-2000; 2000WO-CA000356.  
 XX 05-APR-1999; 99US-0127888P.  
 XX (UUVI-) UNIV VICTORIA.  
 XX White AP, Doran JL, Collison SK, Kay WM;  
 XX WPI; 2000-672631/65.  
 DR N-PSDB; AAC64629.  
 XX Recombinant agfa gene having a segment replaced by a foreign DNA sequence  
 PT which encodes foreign epitope or antigen, expresses recombinant Agfa  
 PT protein useful for eliciting immune response in animal.  
 XX Disclosure; Page 138; 139pp; English.

XX The present invention describes a recombinant agfa gene (1) where a  
 CC segment of the gene has been replaced by a segment of a foreign DNA  
 CC sequence which encodes a foreign epitope or antigen. Also described are:  
 CC (1) use of thin aggregative fimbriae (SEF17/TAf) nucleation depended  
 CC assembly system of strains of Salmonella, Escherichia coli and  
 CC Enterobacteriaceae for the production of fimbriae comprising recombinant  
 CC Agfa, CsgA and Agfa-homologue fimbriae subunits, respectively; (2)  
 CC directing recombination of a recombinant gene into the chromosome of the  
 CC homologous species; (3) directing recombination of a recombinant gene  
 CC back into the chromosome of the homologous species, replacing the native  
 CC copy of that gene; and (4) eliciting an immune response in an animal,  
 CC comprising separating an amino acid polymer comprising a recombinant Agfa  
 CC protein containing a replacement segment or segments of foreign amino  
 CC acid sequence or sequences grown on a Salmonella, E. coli or  
 CC Enterobacteriaceae host cell, from the host cell and introducing the  
 CC polymer into the animal in conjunction with a carrier or diluent. (1) is  
 CC useful for the expression of recombinant Agfa protein which is useful for  
 CC eliciting an immune response in an animal. In a fibrial presentation  
 CC system the heterologous antigens are presented in high numbers (up to  
 CC 500,000 copies/cell), the hybrid fimbriae protein possesses both the  
 CC immunogenicity and adhesion properties relevant for an efficient live  
 CC vaccine, the carrier fimbrial subunit proteins are usually strong  
 CC immunogens, which may be important for directing an immune response  
 CC against the inserted epitope, and hybrid fimbriae are easy and  
 CC inexpensive to purify in large amount. The present sequence is given in  
 CC the exemplification of the present invention  
 XX  
 XX Sequence 151 AA;

Query Match 91.6%; Score 709; DB 3; Length 151;  
 Best Local Similarity 89.9%; Pred. No. 3.7e-64;  
 Matches 143; Conservative 0; Mismatches 0; Indels 16; Gaps 2;  
 QY 1 MKLLKVAAPAAIVVSGSALAGVVPQWGGGNGHNGGNSGPDSTLSIYQYGSANAALALQ 60  
 DB 1 MKLLKVAAPAAIVVSGSALAGVVPQWGGGNGHNGGNSGPDSTLSIYQYGSANAALALQ 57  
 QY 61 SDARKYDQLVTRVVTTHMAHAGGADNSTIETQNGFRNNATIDQNAKNSD 112  
 DB 58 -----YDQLVTRVVTTHMAHAGGADNSTIETQNGFRNNATIDQNAKNSD 112  
 QY 113 ITVGQYGGNNAALVNQTSADSSVMVRQVGFNNATANQY 151  
 DB 113 ITVGQYGGNNAALVNQTSADSSVMVRQVGFNNATANQY 151

RESULT 3  
 AAR74625  
 ID AAR74625 standard; protein; 151 AA.  
 AC AAR74625;  
 XX  
 XX 25-MAR-2003 (revised)  
 DT 26-JUN-1995 (first entry)  
 DE Agfa sequence.  
 XX Salmonella; Agfa; vaccine.  
 XX Salmonella.  
 XX WO9425598-A2.  
 PD 10-NOV-1994.  
 XX 26-APR-1994; 94WO-IB000207.  
 XX 26-APR-1993; 93US-00054452.  
 XX (UUVI-) UNIV VICTORIA INNOVATION & DEV CORP.  
 PA (KING/) KING J.  
 XX

PI Kay WW, Collinson SK, Clouthier SC, Doran JL;  
 XX WPI; 1994-358275/44.  
 DR N-PSDB; AAQ87467.  
 DR  
 PT Eliciting an immune response to Salmonella - using attenuated Salmonella  
 PT strains, vector constructs, or compsns. contg. fimbrial type proteins.  
 XX  
 PS Disclosure; Fig 7B; 95pp; English.  
 XX  
 CC The Salmonella Agfa protein and DNA are used in vaccine and genetic  
 CC immunization compositions, respectively, to elicit an immune response to  
 CC Salmonella in animals (e.g. food producing animals) and humans. (Updated  
 CC on 25-MAR-2003 to correct PN field.)  
 XX  
 SQ Sequence 151 AA;  
 Query Match 89.0%; Score 689; DB 2; Length 151;  
 Best Local Similarity 90.7%; Pred. No. 4.1e-62;  
 Matches 137; Conservative 3; Mismatches 11; Indels 0; Gaps 0;  
 QY 1 M K L L K V A A F A I V V G S S A L A G V V P Q W G G G N H G G N S G P D S T L S I Y Q Y G S A N A A L A Q 60  
 D b 1 M K L L K V A A F A I V V G S S A L A G V V P Q W G G G N H G G N S G P D S T L S I Y Q Y G S A N A A L A Q 60  
 QY 61 S D A R K Y D O L V T R V V T H E M A H A G Q G A D N S T I E L T Q N G F R N N A T I D O W N A K N S D I T V G Q Y G G 120  
 D b 61 S D A R K S E T T I T Q S G Y G N G A D V G Q G A D N S T I E L T Q N G F R N N A T I D O W N A K N S D I T V G Q Y G G 120  
 QY 121 N N A A L V N Q T A S D S S V M V R Q V G F G N N A T A N Q Y 151  
 D b 121 N N A A L V N Q T A S D S S V M V R Q V G F G N N A T A N Q Y 151  
 RESULT 4  
 AAB36341  
 ID AAB36341 standard; protein; 151 AA.  
 AC AAB36341;  
 XX  
 DT 26-FEB-2001 (first entry)  
 XX  
 DE Salmonella enteritidis Agfa amino acid sequence SEQ ID NO:5.  
 XX  
 KW Salmonella; agfa; chromosomal gene replacement; fimbria; epitope;  
 KW vaccine; immune response; immunogen.  
 XX  
 OS Salmonella enteritidis.  
 XX  
 PN WO200060102-A2.  
 XX  
 PD 12-OCT-2000.  
 XX  
 XX 05-APR-2000; 2000WO-CA000356.  
 PF  
 XX 05-APR-1999; 99US-0127888P.  
 PR  
 XX (UYVI-) UNIV VICTORIA.  
 XX  
 XX White AP, Doran JL, Collinson SK, Kay WW;  
 XX  
 XX WPI; 2000-672631/65.  
 XX N-PSDB; AAC64617.  
 DR  
 XX Recombinant agfa gene having a segment replaced by a foreign DNA sequence  
 XX which encodes foreign epitope or antigen, expresses recombinant Agfa  
 XX protein useful for eliciting immune response in animal.  
 PT  
 PT Disclosure; Page 135; 139pp; English.  
 PS  
 XX The present invention describes a recombinant agfa gene (I) where a  
 CC segment of the gene has been replaced by a segment of a foreign DNA  
 CC sequence which encodes a foreign epitope or antigen. Also described are:

(1) use of thin aggregative fimbriae (SEF17/TAF) nucleation depended  
 assembly system of strains of Salmonella, Escherichia coli and  
 Enterobacteriaceae for the production of fimbriae comprising recombinant  
 Agfa, CsgA, and Agfa-homologue fimbria subunits, respectively; (2)  
 directing recombination of a recombinant gene into the chromosome of the  
 homologous species; (3) directing recombination of a recombinant gene  
 back into the chromosome of the homologous species, replacing the native  
 copy of that gene; and (4) eliciting an immune response in an animal,  
 comprising separating an amino acid polymer comprising a recombinant Agfa  
 protein containing a replacement segment or segments of foreign amino  
 acid sequence or sequences grown on a Salmonella, E. coli or  
 Enterobacteriaceae host cell, from the host cell and introducing the  
 polymer into the animal in conjunction with a carrier or diluent. (I) is  
 useful for the expression of recombinant Agfa protein which is useful for  
 eliciting an immune response in an animal. In a fimbrial presentation  
 system the heterologous antigens are presented in high numbers (up to  
 500,000 copies/cell), the hybrid fimbria protein possesses both the  
 immunogenicity and adhesion properties relevant for an efficient live  
 vaccine, the carrier fimbrial subunit proteins are usually strong  
 immunogens, which may be important for directing an immune response  
 against the inserted epitope, and hybrid fimbriae are easy and  
 inexpensive to purify in large amount. The present sequence is given in  
 the exemplification of the present invention  
 Sequence 151 AA;  
 Query Match 89.0%; Score 689; DB 3; Length 151;  
 Best Local Similarity 90.7%; Pred. No. 4.1e-62;  
 Matches 137; Conservative 3; Mismatches 11; Indels 0; Gaps 0;  
 QY 1 M K L L K V A A F A I V V G S S A L A G V V P Q W G G G N H G G N S G P D S T L S I Y Q Y G S A N A A L A Q 60  
 D b 1 M K L L K V A A F A I V V G S S A L A G V V P Q W G G G N H G G N S G P D S T L S I Y Q Y G S A N A A L A Q 60  
 QY 61 S D A R K Y D O L V T R V V T H E M A H A G Q G A D N S T I E L T Q N G F R N N A T I D O W N A K N S D I T V G Q Y G G 120  
 D b 61 S D A R K S E T T I T Q S G Y G N G A D V G Q G A D N S T I E L T Q N G F R N N A T I D O W N A K N S D I T V G Q Y G G 120  
 QY 121 N N A A L V N Q T A S D S S V M V R Q V G F G N N A T A N Q Y 151  
 D b 121 N N A A L V N Q T A S D S S V M V R Q V G F G N N A T A N Q Y 151  
 RESULT 5  
 AAW23570  
 ID AAW23570 standard; protein; 151 AA.  
 AC AAW23570;  
 XX  
 DT 25-MAR-2003 (revised)  
 DT 29-SEP-1997 (first entry)  
 XX  
 DE Salmonella enteritidis 27655-3b agfa.  
 XX  
 KW Enteropathogenic bacteria; enterobacteria; S.enteritidis; antibody.  
 XX  
 OS Salmonella enteritidis.  
 XX  
 PH Key Location/Qualifiers  
 FT Misc-difference 123 /note= "Encoded by GCC"  
 FT  
 XX US5635617-A.  
 XX  
 XX 03-JUN-1997.  
 PD  
 XX 26-APR-1994; 94US-00233788.  
 PF  
 XX 26-APR-1993; 93US-00054452.  
 PR  
 XX (UYVI-) UNIV VICTORIA INNOVATION & DEV CORP.  
 PA  
 XX Collinson SK, Kay WW, Doran JL;  
 PI

XX WPI; 1997-309886/28.  
 DR N-PSDB; AAT74142.  
 XX  
 XX Isolated Salmonella gene agfa - used for diagnosis of Salmonella or  
 PT enteropathogenic bacteria of the Enterobacteriaceae family.  
 XX  
 XX Example 2; Fig 7; 85pp; English.  
 XX  
 XX The present sequence represents agfa encoded by the full agfa gene  
 CC derived from Salmonella enteritidis 27655-3b. The nucleic acid can be  
 CC used to provide diagnostic assays for Salmonella and/or enteropathogenic  
 CC bacteria of the family Enterobacteriaceae. It can also be used to provide  
 CC proteins and antibodies which can be used for assays. The nucleic acid  
 CC sequence can be used to provide probes or primers which can specifically  
 CC hybridise to nucleic acid molecules from greater than 99% of Salmonella  
 CC strains that are pathogenic to warm-blooded animals relative to nucleic  
 CC acid molecules from virtually all other microbial organisms. (Updated on  
 CC 25-MAR-2003 to correct PF field.)  
 XX  
 XX Sequence 151 AA;  
 SQ  
 Query Match 88.4%; Score 684; DB 2; Length 151;  
 Best Local Similarity 90.1%; Pred. NO. 1.3e-61;  
 Matches 136; Conservative 3; Mismatches 12; Indels 0; Gaps 0;  
 QY 1 MKLLKVAFAAIVVSGSALAGVVPQGGGNGHNGGNSGPDSTLSIYQYGSANAALALQ 60  
 DB 1 MKLLKVAFAAIVVSGSALAGVVPQGGGNGHNGGNSGPDSTLSIYQYGSANAALALQ 60  
 QY 61 SDARKYDQLVTRVVTTHMAHAGQADNSTIELTQNGFRNNATIDOWNAKNSDITVGOYGG 120  
 DB 61 SDARKSETTITQSGYNGADVGQADNSTIELTQNGFRNNATIDOWNAKNSDITVGOYGG 120  
 QY 121 NNALVNOTASDSSVWVRQVGFNNATANQY 151  
 DB 121 NNPAVNOTASDSSVWVRQVGFNNATANQY 151  
 RESULT 6  
 AAB36354  
 ID AAB36354 standard; protein; 151 AA.  
 XX  
 AC AAB36354;  
 XX  
 DT 26-FEB-2001 (first entry)  
 XX  
 DE Agfa::PT3#9 amino acid sequence SEQ ID NO:28.  
 KW Salmonella; agfa; chromosomal gene replacement; fimbria; epitope;  
 XX vaccine; immune response; immunogen.  
 OS Salmonella enteritidis.  
 OS Escherichia coli.  
 OS Synthetic.  
 XX  
 FN WO200060102-A2.  
 XX  
 XX 12-OCT-2000.  
 PD  
 XX  
 XX 05-APR-2000; 2000WO-CA000356.  
 PF  
 XX  
 PR 05-APR-1999; 99US-0127888P.  
 XX  
 XX (UYVI-) UNIV VICTORIA.  
 PA  
 XX White AP, Doran JL, Collison SK, Kay Ww;  
 XX WPI; 2000-672631/65.  
 DR N-PSDB; AAC64630.  
 XX  
 XX Recombinant agfa gene having a segment replaced by a foreign DNA sequence  
 PT which encodes foreign epitope or antigen, expresses recombinant Agfa

PT protein useful for eliciting immune response in animal.  
 XX  
 PS Disclosure; Page 138; 139pp; English.  
 XX  
 XX The present invention describes a recombinant agfa gene (I) where a  
 CC segment of the gene has been replaced by a segment of a foreign DNA  
 CC sequence which encodes a foreign epitope or antigen. Also described are:  
 CC (1) use of thin aggregative fimbriae (Sef77/RAF) nucleation depended  
 CC Enterobacteriaceae for the production of fimbriae comprising recombinant  
 CC Agfa, CsgA and Agfa-homologue fimbria subunits, respectively; (2)  
 CC directing recombination of a recombinant gene into the chromosome of the  
 CC homologous species; (3) directing recombination of a recombinant gene  
 CC back into the chromosome of the homologous species, replacing the native  
 CC copy of that gene; and (4) eliciting an immune response in an animal,  
 CC comprising separating an amino acid polymer comprising a recombinant Agfa  
 CC protein containing a replacement segment or segments of foreign amino  
 CC acid sequence or sequences grown on a Salmonella, E. coli or  
 CC Enterobacteriaceae host cell, from the host cell and introducing the  
 CC polymer into the animal in conjunction with a carrier or diluent. (I) is  
 CC useful for the expression of recombinant Agfa protein which is useful for  
 CC eliciting an immune response in an animal. In a fimbrial presentation  
 CC system the heterologous antigens are presented in high numbers (up to  
 CC 500,000 copies/cell), the hybrid fimbria protein possesses both the  
 CC immunogenicity and adhesion properties relevant for an efficient live  
 CC vaccine, the carrier fimbrial subunit proteins are usually strong  
 CC immunogens, which may be important for directing an immune response  
 CC against the inserted epitope, and hybrid fimbriae are easy and  
 CC inexpensive to purify in large amount. The present sequence is given in  
 CC the exemplification of the present invention  
 XX  
 SQ Sequence 151 AA;  
 Query Match 85.5%; Score 662; DB 3; Length 151;  
 Best Local Similarity 81.9%; Pred. No. 2.3e-59;  
 Matches 136; Conservative 0; Mismatches 0; Indels 30; Gaps 2;  
 QY 1 MKLLKVAFAAIVVSGSALAGVVPQGGGNGHNGGNSGPDSTLSIYQYGSANAALALQ 60  
 DB 1 MKLLKVAFAAIVVSGSALAGVVPQGGGNGHNGGNSGPDSTLSIYQYGSANAALALQ 60  
 QY 61 SDARK-----YDQLVTRVVTTHMAHAGQADNSTIELTQNGFRNNATIDQ 105  
 DB 61 SDARKSETTITQSGYNGADYDQLVTRVVTTHMAHA-----FNNATIDQ 105  
 QY 106 WNAKNSDITVGOYGGNNAALVNOTASDSSVWVRQVGFNNATANQY 151  
 DB 106 WNAKNSDITVGOYGGNNAALVNOTASDSSVWVRQVGFNNATANQY 151  
 RESULT 7  
 AAB36349  
 ID AAB36349 standard; protein; 151 AA.  
 XX  
 AC AAB36349;  
 XX  
 DT 26-FEB-2001 (first entry)  
 XX  
 DE Agfa::PT3#4 amino acid sequence SEQ ID NO:18.  
 KW Salmonella; agfa; chromosomal gene replacement; fimbria; epitope;  
 XX vaccine; immune response; immunogen.  
 OS Salmonella enteritidis.  
 OS Escherichia coli.  
 OS Synthetic.  
 XX  
 FN WO200060102-A2.  
 XX  
 XX 12-OCT-2000.  
 PD  
 XX  
 XX 05-APR-2000; 2000WO-CA000356.  
 PF  
 XX



PR 05-APR-1999; 99US-0127888P.  
PA (UYVI-) UNIV VICTORIA.  
PI White AP, Doran JL, Collison SK, Kay WW;  
XX WPI: 2000-672631/65.  
XX N-PSDB; AAC64825.  
XX Recombinant agfa gene having a segment replaced by a foreign DNA sequence  
PT which encodes foreign epitope or antigen, expresses recombinant Agfa  
PT protein useful for eliciting immune response in animal.  
XX Disclosure; Page 136; 139pp; English.  
XX The present invention describes a recombinant agfa gene (I) where a  
CC segment of the gene has been replaced by a segment of a foreign DNA  
CC sequence which encodes a foreign epitope or antigen. Also described are:  
CC (1) use of thin aggregative fimbriae (SEF17/TAf) nucleation depended  
CC assembly system of strains of Salmonella, Escherichia coli and  
CC Enterobacteriaceae for the production of fimbriae comprising recombinant  
CC Agfa, CsgA and Agfa-homologue fimbrial subunits, respectively; (2)  
CC directing recombination of a recombinant gene into the chromosome of the  
CC homologous species; (3) directing recombination of a recombinant gene  
CC back into the chromosome of the homologous species, replacing the native  
CC copy of that gene; and (4) eliciting an immune response in an animal,  
CC comprising separating an amino acid polymer comprising a recombinant Agfa  
CC protein containing a replacement segment or segments of foreign amino  
CC acid sequence or sequences grown on a Salmonella, E. coli or  
CC Enterobacteriaceae host cell, from the host cell and introducing the  
CC polymer into the animal in conjunction with a carrier or diluent. (I) is  
CC useful for the expression of recombinant Agfa protein which is useful for  
CC eliciting an immune response in an animal. In a fimbrial presentation  
CC system the heterologous antigens are presented in high numbers (up to  
CC 500,000 copies/cell) the hybrid fimbrial protein possesses both the  
CC immunogenicity and adhesion properties relevant for an efficient live  
CC vaccine, the carrier fimbrial subunit proteins are usually strong  
CC immunogens, which may be important for directing an immune response  
CC against the inserted epitope, and hybrid fimbriae are easy and  
CC inexpensive to purify in large amount. The present sequence is given in  
CC the exemplification of the present invention  
XX Sequence 151 AA;  
XX Query Match 80.4%; Score 622; DB 3; Length 151;  
XX Best Local Similarity 73.6%; Pred. No. 2.8e-55;  
XX Matches 128; Conservative 0; Mismatches 0; Indels 46; Gaps 2;  
QY 1 MKLLKVAAPAAIVVSGSALAGVVPQWGGGNGHNGGNSGPDSTLSIYQYGSANAALALQ 60  
DB 1 MKLLKVAAPAAIVVSGSALAGVVPQWGGGNGHNGGNSGPD----- 42  
QY 61 SDARKYDQLVTRVVTHEMAHA-----GQADNSTIELTQNGF 97  
DB 43 -----YDQLVTRVVTHEMAHALQSDARKSETTITQSGYNGADVQGGADNSTIELTQNGF 97  
QY 98 RNNATIDQWNAKNSDITVGYGGNNAALVNOTASDSVMVRQVGFGNNTATNOY 151  
DB 98 RNNATIDQWNAKNSDITVGYGGNNAALVNOTASDSVMVRQVGFGNNTATNOY 151  
RESULT 8  
ID AAB36351 standard; protein; 151 AA.  
XX AAB36351;  
XX 26-FEB-2001. (first entry)  
DE Agfa: P3#6 amino acid sequence SEQ ID NO.22.  
XX Salmonella; agfa; chromosomal gene replacement; fimbrial; epitope;  
KW vaccine; immune response; immunogen.

XX Salmonella enteritidis.  
OS Escherichia coli.  
OS Synthetic.  
PN WO200060102-A2.  
XX 12-OCT-2000.  
XX 05-APR-2000; 2000WO-CA000356.  
XX 05-APR-1999; 99US-0127888P.  
XX (UYVI-) UNIV VICTORIA.  
XX White AP, Doran JL, Collison SK, Kay WW;  
PI WPI: 2000-672631/65.  
XX N-PSDB; AAC64627.  
XX Recombinant agfa gene having a segment replaced by a foreign DNA sequence  
PT which encodes foreign epitope or antigen, expresses recombinant Agfa  
PT protein useful for eliciting immune response in animal.  
XX Disclosure; Page 137; 139pp; English.  
XX The present invention describes a recombinant agfa gene (I) where a  
CC segment of the gene has been replaced by a segment of a foreign DNA  
CC sequence which encodes a foreign epitope or antigen. Also described are:  
CC (1) use of thin aggregative fimbriae (SEF17/TAf) nucleation depended  
CC assembly system of strains of Salmonella, Escherichia coli and  
CC Enterobacteriaceae for the production of fimbriae comprising recombinant  
CC Agfa, CsgA and Agfa-homologue fimbrial subunits, respectively; (2)  
CC directing recombination of a recombinant gene into the chromosome of the  
CC homologous species; (3) directing recombination of a recombinant gene  
CC back into the chromosome of the homologous species, replacing the native  
CC copy of that gene; and (4) eliciting an immune response in an animal,  
CC comprising separating an amino acid polymer comprising a recombinant Agfa  
CC protein containing a replacement segment or segments of foreign amino  
CC acid sequence or sequences grown on a Salmonella, E. coli or  
CC Enterobacteriaceae host cell, from the host cell and introducing the  
CC polymer into the animal in conjunction with a carrier or diluent. (I) is  
CC useful for the expression of recombinant Agfa protein which is useful for  
CC eliciting an immune response in an animal. In a fimbrial presentation  
CC system the heterologous antigens are presented in high numbers (up to  
CC 500,000 copies/cell), the hybrid fimbrial protein possesses both the  
CC immunogenicity and adhesion properties relevant for an efficient live  
CC vaccine, the carrier fimbrial subunit proteins are usually strong  
CC immunogens, which may be important for directing an immune response  
CC against the inserted epitope, and hybrid fimbriae are easy and  
CC inexpensive to purify in large amount. The present sequence is given in  
CC the exemplification of the present invention  
XX Sequence 151 AA;  
XX Query Match 80.0%; Score 619; DB 3; Length 151;  
XX Best Local Similarity 74.6%; Pred. No. 5.6e-55;  
XX Matches 129; Conservative 0; Mismatches 0; Indels 44; Gaps 2;  
QY 1 MKLLKVAAPAAIVVSGSALAGVVPQWGGGNGHNGGNSGPDSTLSIYQYGSANAALALQ 60  
DB 1 MKLLKVAAPAAIVVSGSALAGVVPQWGGGNGHNGGNSGPDSTLSIYQYGSANAALALQ 60  
QY 61 SDARK-----YDQLVTRVVTHEMAHAGQADNSTIELTQNGF 98  
DB 61 SDARKSETTITQSGYNGADVQGGADNYDQLVTRVVTHEMAHA----- 103  
QY 99 RNNATIDQWNAKNSDITVGYGGNNAALVNOTASDSVMVRQVGFGNNTATNOY 151  
DB 104 -----DOWNAKNSDITVGYGGNNAALVNOTASDSVMVRQVGFGNNTATNOY 151  
RESULT 9

AAB36346  
 ID AAB36346 standard; protein; 151 AA.  
 AC AAB36346;  
 XX  
 DT 26-FEB-2001 (first entry)  
 XX  
 DE Agfa::PT3#1 amino acid sequence SEQ ID NO:12.  
 XX  
 KW Salmonella; agfa; chromosomal gene replacement; fimbrin; epitope;  
 KW vaccine; immune response; immunogen.  
 XX  
 OS Salmonella enteritidis.  
 OS Escherichia coli.  
 OS Synthetic.  
 XX  
 FN WO200060102-A2.  
 XX  
 PD 12-OCT-2000.  
 XX  
 PF 05-APR-2000; 2000WO-CA000356.  
 XX  
 PR 05-APR-1999; 99US-0127888P.  
 XX  
 PA (UYVI-) UNIV VICTORIA.  
 XX  
 PI White AP, Doran JL, Collison SK, Kay WW;  
 XX  
 DR WPI; 2000-672631/65.  
 DR N-PSDB; AAC64622.  
 XX  
 PT Recombinant agfa gene having a segment replaced by a foreign DNA sequence  
 PT which encodes foreign epitope or antigen, expresses recombinant Agfa  
 PT protein useful for eliciting immune response in animal.  
 XX  
 PS Disclosure; Page 135; 139pp; English.  
 XX  
 CC The present invention describes a recombinant agfa gene (1) where a  
 CC segment of the gene has been replaced by a segment of a foreign DNA  
 CC sequence which encodes a foreign epitope or antigen. Also described are:  
 CC (1) use of thin aggregative fimbriae (SEF17/TAF) nucleation depended  
 CC assembly system of strains of Salmonella, Escherichia coli and  
 CC Enterobacteriaceae for the production of fimbriae comprising recombinant  
 CC Agfa, CsgA and Agfa-homologue fimbrial subunits, respectively; (2)  
 CC directing recombination of a recombinant gene into the chromosome of the  
 CC homologous species; (3) directing recombination of a recombinant gene  
 CC back into the chromosome of the homologous species, replacing the native  
 CC copy of that gene; and (4) eliciting an immune response in an animal,  
 CC comprising separating an amino acid polymer comprising a recombinant Agfa  
 CC protein containing a replacement segment or segments of foreign amino  
 CC acid sequence or sequences grown on a Salmonella, E. coli or  
 CC Enterobacteriaceae host cell, from the host cell and introducing the  
 CC polymer into the animal in conjunction with a carrier or diluent. (1) is  
 CC useful for the expression of recombinant Agfa protein which is useful for  
 CC eliciting an immune response in an animal. In a fimbrial presentation  
 CC system the heterologous antigens are presented in high numbers (up to  
 CC 500,000 copies/cell), the hybrid fimbrial protein possesses both the  
 CC immunogenicity and adhesion properties relevant for an efficient live  
 CC vaccine, the carrier fimbrial subunit proteins are usually strong  
 CC immunogens, which may be important for directing an immune response  
 CC against the inserted epitope, and hybrid fimbriae are easy and  
 CC inexpensive to purify in large amount. The present sequence is given in  
 CC the exemplification of the present invention  
 XX  
 SQ Sequence 151 AA;  
 Query Match 78.9%; Score 611; DB 3; Length 151;  
 Best Local Similarity 80.8%; Pred. No. 3.7e-54;  
 Matches 122; Conservative 7; Mismatches 22; Indels 0; Gaps 0;  
 1 MKLLKVAFAAIVVSGSALAGVVPQGGGNGGNGSGSDSTSIYQGSANALALQ 60  
 1 MKLLKVAFAAIVVSGSALAGVVPQGGGNGGNGSGSDSTSIYQGSANALALQ 60

QY 61 SDARKYDQLVTRVVTHEMAHAGGADNSTIELTONGFRNNATIDOWNAKNSDITVGOYGG 120  
 DB 61 SDARKSETTITQSGYNGADVGGADNSTIELTONGFRNNATIDOWNAKNSDITVGOYGG 120  
 XX  
 QY 121 NNAALVNQATSDSSVMVRQVGFNNATANCY 151  
 DB 121 NNAALVNYDQLVTRVVTHEMAHANNATANCY 151  
 XX  
 RESULT 10  
 AAB36347  
 ID AAB36347 standard; protein; 151 AA.  
 AC AAB36347;  
 XX  
 DT 26-FEB-2001 (first entry)  
 XX  
 DE Agfa::PT3#2 amino acid sequence SEQ ID NO:14.  
 XX  
 KW Salmonella; agfa; chromosomal gene replacement; fimbrin; epitope;  
 KW vaccine; immune response; immunogen.  
 XX  
 OS Salmonella enteritidis.  
 OS Escherichia coli.  
 OS Synthetic.  
 XX  
 FN WO200060102-A2.  
 XX  
 PD 12-OCT-2000.  
 XX  
 PF 05-APR-2000; 2000WO-CA000356.  
 XX  
 PR 05-APR-1999; 99US-0127888P.  
 XX  
 PA (UYVI-) UNIV VICTORIA.  
 XX  
 PI White AP, Doran JL, Collison SK, Kay WW;  
 XX  
 DR WPI; 2000-672631/65.  
 DR N-PSDB; AAC64623.  
 XX  
 PT Recombinant agfa gene having a segment replaced by a foreign DNA sequence  
 PT which encodes foreign epitope or antigen, expresses recombinant Agfa  
 PT protein useful for eliciting immune response in animal.  
 XX  
 PS Disclosure; Page 136; 139pp; English.  
 XX  
 CC The present invention describes a recombinant agfa gene (1) where a  
 CC segment of the gene has been replaced by a segment of a foreign DNA  
 CC sequence which encodes a foreign epitope or antigen. Also described are:  
 CC (1) use of thin aggregative fimbriae (SEF17/TAF) nucleation depended  
 CC assembly system of strains of Salmonella, Escherichia coli and  
 CC Enterobacteriaceae for the production of fimbriae comprising recombinant  
 CC Agfa, CsgA and Agfa-homologue fimbrial subunits, respectively; (2)  
 CC directing recombination of a recombinant gene into the chromosome of the  
 CC homologous species; (3) directing recombination of a recombinant gene  
 CC back into the chromosome of the homologous species, replacing the native  
 CC copy of that gene; and (4) eliciting an immune response in an animal,  
 CC comprising separating an amino acid polymer comprising a recombinant Agfa  
 CC protein containing a replacement segment or segments of foreign amino  
 CC acid sequence or sequences grown on a Salmonella, E. coli or  
 CC Enterobacteriaceae host cell, from the host cell and introducing the  
 CC polymer into the animal in conjunction with a carrier or diluent. (1) is  
 CC useful for the expression of recombinant Agfa protein which is useful for  
 CC eliciting an immune response in an animal. In a fimbrial presentation  
 CC system the heterologous antigens are presented in high numbers (up to  
 CC 500,000 copies/cell), the hybrid fimbrial protein possesses both the  
 CC immunogenicity and adhesion properties relevant for an efficient live  
 CC vaccine, the carrier fimbrial subunit proteins are usually strong  
 CC immunogens, which may be important for directing an immune response  
 CC against the inserted epitope, and hybrid fimbriae are easy and  
 CC inexpensive to purify in large amount. The present sequence is given in  
 CC the exemplification of the present invention

CC the exemplification of the present invention

XX Sequence 151 AA;

Query Match 78.7%; Score 609; DB 3; Length 151;

Best Local Similarity 81.5%; Pred. No. 5.8e-54;

Matches 123; Conservative 6; Mismatches 22; Indels 0; Gaps 0;

QY 1 MKLLKVAFAAIVVSGSALAGVVPQWGGGNGHNGSGPDSLTLSIYQGSANAALALQ 60

DB 1 MKLLKVAFAAIVVSGSALAGVVPQWGGGNGHNGSGPDSLTLSIYQGSANAALALQ 60

QY 61 SDARKYDQLVTRVTHEMAHAGGADNSTIELTQNGFRNNATIDQWNAKNSDITVGYGG 120

DB 61 SDARKSETTITQSGYNGADVGGADNSTIELTQNGFRNNATIDQWNAKNSDITVGYDDQ 120

QY 121 NNAALVNOTASDSSVMVRQVGFNNATANQY 151

DB 121 LVTRVTHEMAHASVMVRQVGFNNATANQY 151

RESULT 11

AAB36352

ID AAB36352 standard; protein; 151 AA.

XX AAB36352;

AC AAB36352;

DT 26-FEB-2001 (first entry)

DE Agfa::PT3#7 amino acid sequence SEQ ID NO:24.

XX Salmonella; agfa; chromosomal gene replacement; fimbrin; epitope;

KW vaccine; immune response; immunogen.

XX Salmonella enteritidis.

OS Escherichia coli.

OS Synthetic.

XX WO200060102-A2.

XX 12-OCT-2000.

XX 05-APR-2000; 2000WO-CA000356.

XX 05-APR-1999; 99US-0127888P.

XX (UYVI-) UNIV VICTORIA.

XX White AP, Doran JL, Collison SK, Kay WW;

XX WPI; 2000-672631/65.

XX N-PSDB; AAC64628.

XX Recombinant agfa gene having a segment replaced by a foreign DNA sequence

PT which encodes foreign epitope or antigen, expresses recombinant Agfa

PT protein useful for eliciting immune response in animal.

XX Disclosure; Page 138; 139pp; English.

XX The present invention describes a recombinant agfa gene (I) where a

CC segment of the gene has been replaced by a segment of a foreign DNA

CC sequence which encodes a foreign epitope or antigen. Also described are:

CC (1) use of thin aggregative fimbriae (SEF77/TAF) nucleation depended

CC assembly system of strains of Salmonella, Escherichia coli and

CC Enterobacteriaceae for the production of fimbriae comprising recombinant

CC Agfa, CsgA and Agfa-homologue fimbrin subunits, respectively; (2)

CC directing recombination of a recombinant gene into the chromosome of the

CC homologous species; (3) directing recombination of a recombinant gene

CC back into the chromosome of the homologous species, replacing the native

CC copy of that gene; and (4) eliciting an immune response in an animal.

CC comprising separating an amino acid polymer comprising a recombinant Agfa

CC protein containing a replacement segment or segments of foreign amino

CC acid sequence or sequences grown on a Salmonella, E. coli or

CC Enterobacteriaceae host cell, from the host cell and introducing the

CC polymer into the animal in conjunction with a carrier or diluent. (I) is

CC useful for the expression of recombinant Agfa protein which is useful for

CC eliciting an immune response in an animal. In a fimbrial presentation

CC system the heterologous antigens are presented in high numbers (up to

CC 500,000 copies/cell), the hybrid fimbrin protein possesses both the

CC immunogenicity and adhesion properties relevant for an efficient live

CC vaccine, the carrier fimbrial subunit proteins are usually strong

CC immunogens, which may be important for directing an immune response

CC against the inserted epitope, and hybrid fimbriae are easy and

CC inexpensive to purify in large amount. The present sequence is given in

CC the exemplification of the present invention

XX Sequence 151 AA;

Query Match 78.3%; Score 606; DB 3; Length 151;

Best Local Similarity 82.1%; Pred. No. 1.2e-53;

Matches 124; Conservative 5; Mismatches 22; Indels 0; Gaps 0;

QY 1 MKLLKVAFAAIVVSGSALAGVVPQWGGGNGHNGSGPDSLTLSIYQGSANAALALQ 60

DB 1 MKLLKVAFAAIVVSGSALAGVVPQWGGGNGHNGSGPDSLTLSIYQGSANAALALQ 60

QY 61 SDARKYDQLVTRVTHEMAHAGGADNSTIELTQNGFRNNATIDQWNAKNSDITVGYGG 120

DB 61 SDARKSETTITQSGYNGADVGGADNSTIELTQNGFRNNATIDQWNAKNSDITVTRVT 120

QY 121 NNAALVNOTASDSSVMVRQVGFNNATANQY 151

DB 121 HEMAHANQTASDSSVMVRQVGFNNATANQY 151

RESULT 12

AAB36355

ID AAB36355 standard; protein; 151 AA.

XX AAB36355;

AC AAB36355;

DT 26-FEB-2001 (first entry)

DE Agfa::PT3#10 amino acid sequence SEQ ID NO:30.

XX Salmonella; agfa; chromosomal gene replacement; fimbrin; epitope;

KW vaccine; immune response; immunogen.

XX Salmonella enteritidis.

OS Escherichia coli.

OS Synthetic.

XX WO200060102-A2.

XX 12-OCT-2000.

XX 05-APR-2000; 2000WO-CA000356.

XX 05-APR-1999; 99US-0127888P.

XX (UYVI-) UNIV VICTORIA.

XX White AP, Doran JL, Collison SK, Kay WW;

XX WPI; 2000-672631/65.

XX N-PSDB; AAC64631.

XX Recombinant agfa gene having a segment replaced by a foreign DNA sequence

PT which encodes foreign epitope or antigen, expresses recombinant Agfa

PT protein useful for eliciting immune response in animal.

XX Disclosure; Page 139; 139pp; English.

XX The present invention describes a recombinant agfa gene (I) where a

CC segment of the gene has been replaced by a segment of a foreign DNA

CC sequence which encodes a foreign epitope or antigen. Also described are:

CC (1) use of thin aggregative fimbriae (SEF17/TAF) nucleation depended  
 CC assembly system of strains of Salmonella, Escherichia coli and  
 CC Enterobacteriaceae for the production of fimbriae comprising recombinant  
 CC AgfA, CsgA and AgfA-homologue fimbrial subunits, respectively; (2)  
 CC directing recombination of a recombinant gene into the chromosome of the  
 CC homologous species; (3) directing recombination of a recombinant gene  
 CC back into the chromosome of the homologous species, replacing the native  
 CC copy of that gene; and (4) eliciting an immune response in an animal,  
 CC comprising separating an amino acid polymer comprising a recombinant AgfA  
 CC protein containing a replacement segment or segments of foreign amino  
 CC acid sequence or sequences grown on a Salmonella, E. coli or  
 CC Enterobacteriaceae host cell, from the host cell and introducing the  
 CC polymer into the animal in conjunction with a carrier or diluent. (1) is  
 CC useful for the expression of recombinant AgfA protein which is useful for  
 CC eliciting an immune response in an animal. In a fimbrial presentation  
 CC system the heterologous antigens are presented in high numbers (up to  
 CC 500,000 copies/cell), the hybrid fimbrial protein possesses both the  
 CC immunogenicity and adhesion properties relevant for an efficient live  
 CC vaccine, the carrier fimbrial subunit proteins are usually strong  
 CC immunogens, which may be important for directing an immune response  
 CC against the inserted epitope, and hybrid fimbriae are easy and  
 CC inexpensive to purify in large amount. The present sequence is given in  
 CC the exemplification of the present invention

XX Sequence 151 AA;

Query Match 77.1%; Score 597; DB 3; Length 151;  
 Best Local Similarity 81.5%; Pred. No. 9.8e-53;  
 Matches 123; Conservative 5; Mismatches 23; Indels 0; Gaps 0;  
 QY 1 MKLLKVAAPAAIIVVSGSALAGVVPQWGGGNGHNGGNSGPDSTLSIYQYGSANAALALQ 60  
 Db 1 MKLLKVAAPAAIIVVSGSALAGVVPQWGGGNGHNGGNSGPDSTLSIYQYGSANAALALQ 60  
 QY 61 SDARKYDQLVTRVVTTHMAHAGGADNSTIELTQNGFRNNATIDOWNAKNSDITVQYGG 120  
 Db 61 SDARKSETTITQSGYNGADVGQADNSTIELTQNGFRNNATIDVLTTRVVTTHMAHAGG 120  
 QY 121 NNAALVNQTSASDSSVMVRQVGFNNATANQY 151  
 Db 121 NNAALVNQTSASDSSVMVRQVGFNNATANQY 151

# RESULT 13

AAB36348  
 ID AAB36348 standard; protein; 151 AA.  
 XX  
 AC AAB36348;  
 DT 26-FEB-2001 (first entry)  
 DE AgfA::PT3#3 amino acid sequence SEQ ID NO:16.  
 XX Salmonella; agfA; chromosomal gene replacement; fimbria; epitope;  
 KW vaccine; immune response; immunogen.  
 XX Salmonella enteritidis.  
 OS Escherichia coli.  
 OS Synthetic.  
 XX  
 FN WO2000060102-A2.  
 PD 12-OCT-2000.  
 XX  
 PF 05-APR-2000; 2000WO-CA000356.  
 XX  
 PR 05-APR-1999; 99US-0127888P.  
 XX  
 PA (UYVI-) UNIV VICTORIA.  
 PI White AP, Doran JL, Collison SK, Kay WW;  
 XX WPI; 2000-672631/65.

DR N-PSDB; AAC64624.

XX Recombinant agfA gene having a segment replaced by a foreign DNA sequence  
 PT which encodes foreign epitope or antigen, expresses recombinant AgfA  
 PT protein useful for eliciting immune response in animal.  
 XX  
 XX Disclosure; Page 136; 139pp; English.

XX The present invention describes a recombinant agfA gene (1) where a  
 CC segment of the gene has been replaced by a segment of a foreign DNA  
 CC sequence which encodes a foreign epitope or antigen. Also described are:  
 CC (1) use of thin aggregative fimbriae (SEF17/TAF) nucleation depended  
 CC assembly system of strains of Salmonella, Escherichia coli and  
 CC Enterobacteriaceae for the production of fimbriae comprising recombinant  
 CC AgfA, CsgA and AgfA-homologue fimbrial subunits, respectively; (2)  
 CC directing recombination of a recombinant gene into the chromosome of the  
 CC homologous species; (3) directing recombination of a recombinant gene  
 CC back into the chromosome of the homologous species; replacing the native  
 CC copy of that gene; and (4) eliciting an immune response in an animal,  
 CC comprising separating an amino acid polymer comprising a recombinant AgfA  
 CC protein containing a replacement segment or segments of foreign amino  
 CC acid sequence or sequences grown on a Salmonella, E. coli or  
 CC Enterobacteriaceae host cell, from the host cell and introducing the  
 CC polymer into the animal in conjunction with a carrier or diluent. (1) is  
 CC useful for the expression of recombinant AgfA protein which is useful for  
 CC eliciting an immune response in an animal. In a fimbrial presentation  
 CC system the heterologous antigens are presented in high numbers (up to  
 CC 500,000 copies/cell), the hybrid fimbrial protein possesses both the  
 CC immunogenicity and adhesion properties relevant for an efficient live  
 CC vaccine, the carrier fimbrial subunit proteins are usually strong  
 CC immunogens, which may be important for directing an immune response  
 CC against the inserted epitope, and hybrid fimbriae are easy and  
 CC inexpensive to purify in large amount. The present sequence is given in  
 CC the exemplification of the present invention

XX Sequence 151 AA;

Query Match 74.2%; Score 574; DB 3; Length 151;  
 Best Local Similarity 80.8%; Pred. No. 2.2e-50;  
 Matches 122; Conservative 6; Mismatches 23; Indels 0; Gaps 0;  
 QY 1 MKLLKVAAPAAIIVVSGSALAGVVPQWGGGNGHNGGNSGPDSTLSIYQYGSANAALALQ 60  
 Db 1 MKLLKVAAPAAIIVVSGSALAGVVDQLVTRVVTTHMAHAGGPDSTLSIYQYGSANAALALQ 60  
 QY 61 SDARKYDQLVTRVVTTHMAHAGGADNSTIELTQNGFRNNATIDOWNAKNSDITVQYGG 120  
 Db 61 SDARKSETTITQSGYNGADVGQADNSTIELTQNGFRNNATIDOWNAKNSDITVQYGG 120  
 QY 121 NNAALVNQTSASDSSVMVRQVGFNNATANQY 151  
 Db 121 NNAALVNQTSASDSSVMVRQVGFNNATANQY 151

# RESULT 14

AAB36343  
 ID AAB36343 standard; protein; 151 AA.  
 XX  
 AC AAB36343;  
 DT 26-FEB-2001 (first entry)  
 DE Escherichia coli CsgA amino acid sequence SEQ ID NO:7.  
 XX Salmonella; agfA; chromosomal gene replacement; fimbria; epitope;  
 KW vaccine; immune response; immunogen.  
 XX Escherichia coli.  
 OS  
 XX WO2000060102-A2.  
 FN  
 PD 12-OCT-2000.  
 XX

PF 05-APR-2000; 2000WO-CA000356.  
XX  
PR 05-APR-1999; 99US-0127888P.  
XX  
FA (UYVI-) UNIV VICTORIA.  
XX  
PI White AP, Doran JL, Collison SK, Kay WW;  
XX  
DR WPI; 2000-672631/65.  
DR N-PSDB; AAC64619.  
XX  
XX Recombinant agfa gene having a segment replaced by a foreign DNA sequence  
PT which encodes foreign epitope or antigen, expresses recombinant Agfa  
PT protein useful for eliciting immune response in animal.  
XX  
PS Disclosure; Page 135; 139pp; English.  
XX  
XX The present invention describes a recombinant agfa gene (I) where a  
CC segment of the gene has been replaced by a segment of a foreign DNA  
CC sequence which encodes a foreign epitope or antigen. Also described are:  
CC (1) use of thin aggregative fimbriae (SEF17/TAF) nucleation depended  
CC assembly system of strains of Salmonella, Escherichia coli and  
CC Enterobacteriaceae for the production of fimbriae comprising recombinant  
CC Agfa, CsgA and Agfa-homologue fimbrial subunits, respectively; (2)  
CC directing recombination of a recombinant gene into the chromosome of the  
CC homologous species; (3) directing recombination of a recombinant gene  
CC back into the chromosome of the homologous species, replacing the native  
CC copy of that gene; and (4) eliciting an immune response in an animal,  
CC comprising separating an amino acid polymer comprising a recombinant Agfa  
CC protein containing a replacement segment or segments of foreign amino  
CC acid sequence or sequences grown on a Salmonella, E. coli or  
CC Enterobacteriaceae host cell, from the host cell and introducing the  
CC polymer into the animal in conjunction with a carrier or diluent. (I) is  
CC useful for the expression of recombinant Agfa protein which is useful for  
CC eliciting an immune response in an animal. In a fimbrial presentation  
CC system the heterologous antigens are presented in high numbers (up to  
CC 500,000 copies/cell), the hybrid fimbrial protein possesses both the  
CC immunogenicity and adhesion properties relevant for an efficient live  
CC vaccine, the carrier fimbrial subunit proteins are usually strong  
CC immunogens, which may be important for directing an immune response  
CC against the inserted epitope, and hybrid fimbriae are easy and  
CC inexpensive to purify in large amount. The present sequence is given in  
CC the exemplification of the present invention  
XX  
XX Sequence 151 AA;  
SQ  
Query Match 68.2%; Score 528; DB 3; Length 151;  
Best Local Similarity 68.9%; Pred. No. 1.1e-45;  
Matches 104; Conservative 20; Mismatches 27; Indels 0; Gaps 0;  
QY 1 MKLLKVAFAAIVVSGSALAGVVPWGCGGNHGGGNSGPDSTLSIYQVGSANAALQ 60  
DB 1 MKLLKVAFAAIVFSGSALAGVVPYGGGNGHGGGNSGPNSELNIYQVGGNSALQ 60  
QY 61 SDARKYDQLVTRVVTHEMAHAGQADNSTIELTQNGFNNATIDOWNAKNSDITVQYGG 120  
DB 61 TDARNSDLTITQHGCGGNGADVQGGSDSSIDLITQRFNGSATLDOWNKNSMTVKQFGG 120  
QY 121 NNAALVNOTASDSSVMVTVQVGFNNATANY 151  
DB 121 GNGAAVDQTASNSVNVTVQVGFNNATAHQY 151  
RESULT 15  
ID ABR82651  
XX ABR82651 standard; protein; 151 AA.  
XX AC ABR82651;  
XX  
DT 04-DEC-2003 (first entry)  
XX  
XX E. coli CsgA subunit 15 kDa protein.

KW Plasma protein; immune response; antibacterial; vaccine; gene therapy.  
XX Escherichia coli.  
XX OS  
XX WO2003064446-A2.  
XX  
XX 07-AUG-2003.  
XX  
XX 30-JAN-2003; 2003WO-EP000943.  
XX  
XX 31-JAN-2002; 2002GB-00002275.  
XX  
XX (HANS-) HANSA MEDICAL RES AB.  
XX  
XX Bjoerck L, Olsen A, Wikstroem M, Herwald H;  
XX  
XX WPI; 2003-646136/61.  
XX N-PSDB; ACF36153.  
XX  
XX New isolated peptide capable of binding a mammalian plasma protein,  
PT useful in the manufacture of a medicament for the prevention and/or  
PT treatment of a bacterial infection, such as Escherichia coli, Salmonella  
PT or Shigella infections.  
XX  
PS Disclosure; Page 41-42; 42pp; English.  
XX  
XX The invention relates to an isolated peptide capable of binding a  
CC mammalian plasma protein or of generating an immune response in a mammal  
CC selected from sequences shown in ABR82642, ABR82648-49. The peptide or  
CC antibody is useful for treating a bacterial infection in a human or  
CC animal or in the manufacture of a medicament for the prophylactic  
CC treatment of a bacterial infection, such as Escherichia coli, Salmonella  
CC or Shigella infection. The peptide that is immobilized on a solid support  
CC is also useful as a reagent for determining the ability of a plasma  
CC protein to bind to bacteria. The present sequence represents an E. coli  
CC 15 kDa protein  
XX  
XX Sequence 151 AA;  
SQ  
Query Match 67.6%; Score 523; DB 7; Length 151;  
Best Local Similarity 68.2%; Pred. No. 3.4e-45;  
Matches 103; Conservative 20; Mismatches 28; Indels 0; Gaps 0;  
QY 1 MKLLKVAFAAIVVSGSALAGVVPWGCGGNHGGGNSGPDSTLSIYQVGSANAALQ 60  
DB 1 MKLLKVAFAAIVFSGSALAGVVPYGGGNGHGGGNSGPNSELNIYQVGGNSALQ 60  
QY 61 SDARKYDQLVTRVVTHEMAHAGQADNSTIELTQNGFNNATIDOWNAKNSDITVQYGG 120  
DB 61 TDARNSDLTITQHGCGGNGADVQGGSDSSIDLITQRFNGSATLDOWNKNSMTVKQFGG 120  
QY 121 NNAALVNOTASDSSVMVTVQVGFNNATANY 151  
DB 121 GNGAAVDQTASNSVNVTVQVGFNNATAHQY 151  
Search completed: August 2, 2004, 14:48:26  
Job time : 45.9 secs



GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: August 2, 2004, 14:40:48 ; Search time 12 Seconds  
(without alignments)  
649.627 Million cell updates/sec

Title: US-09-543-407-20

Perfect score: 774

Sequence: 1 MKLLKVAFAAIVVSGSALA.....DSSVMVRQVFGNATANY 151

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 389414 seqs, 51625971 residues

Total number of hits satisfying chosen parameters: 389414

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

Issued Patents AA:\*

1: /cgn2\_6/ptodata/2/iaa/5A COMB.pap.\*

2: /cgn2\_6/ptodata/2/iaa/5B COMB.pap.\*

3: /cgn2\_6/ptodata/2/iaa/6A COMB.pap.\*

4: /cgn2\_6/ptodata/2/iaa/6B COMB.pap.\*

5: /cgn2\_6/ptodata/2/iaa/PCTUS COMB.pap.\*

6: /cgn2\_6/ptodata/2/iaa/backfiles1.pap.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	684	88.4	151	1	US-08-233-788A-59
2	504	65.1	120	1	US-08-233-788A-57
3	83.5	10.8	208	4	US-09-552-991A-27661
4	83.5	10.8	738	3	US-08-864-038A-3
5	79.5	10.3	321	4	US-09-498-520A-18
6	79.5	10.3	975	4	US-09-328-352-4764
7	78	10.1	363	1	US-08-458-023B-6
8	78	10.1	435	2	US-08-331-515A-2
9	78	10.1	435	3	US-09-168-406A-2
10	77.5	10.0	254	3	US-09-128-450-26
11	77.5	10.0	254	4	US-09-323-494-26
12	77.5	10.0	364	1	US-07-792-259-17
13	77	9.9	943	4	US-09-056-556-204
14	77	9.9	943	4	US-09-072-596-199
15	77	9.9	943	4	US-09-477-135A-131
16	77	9.9	943	4	US-09-072-967-204
17	76.5	9.9	1739	4	US-09-540-236-3739
18	76	9.8	252	4	US-09-431-887-32
19	76	9.8	273	4	US-09-328-352-6167
20	76	9.8	892	4	US-09-336-447A-5
21	76	9.8	2123	3	US-08-968-685A-10
22	75.5	9.8	232	4	US-09-555-352-10
23	75.5	9.8	254	4	US-09-431-887-23
24	75.5	9.8	415	4	US-09-025-769B-280
25	75	9.7	208	3	US-09-128-450-18
26	75	9.7	208	4	US-08-823-494-18
27	75	9.7	349	4	US-09-300-971A-9

28	74.5	9.6	304	1	US-07-851-976B-8	Sequence 8, Appli
29	74.5	9.6	304	1	US-08-251-609-8	Sequence 8, Appli
30	74.5	9.6	304	1	US-08-401-136-8	Sequence 8, Appli
31	74.5	9.6	304	3	US-08-850-554-8	Sequence 8, Appli
32	73.5	9.5	211	1	US-08-276-852-34	Sequence 34, Appl
33	73.5	9.5	211	1	US-08-133-011-16	Sequence 16, Appl
34	73.5	9.5	211	1	US-08-322-730A-16	Sequence 16, Appl
35	73.5	9.5	211	1	US-08-387-874-16	Sequence 16, Appl
36	73.5	9.5	211	1	US-08-899-575-34	Sequence 34, Appl
37	73.5	9.5	211	1	US-08-899-575-34	Sequence 34, Appl
38	73.5	9.5	211	2	US-08-383-619-16	Sequence 16, Appl
39	73.5	9.5	211	3	US-08-907-739-16	Sequence 16, Appl
40	73.5	9.5	211	4	US-09-729-597-16	Sequence 16, Appl
41	73.5	9.5	211	5	PCT-US93-08364-16	Sequence 16, Appl
42	73.5	9.5	211	5	PCT-US95-08743-34	Sequence 34, Appl
43	73.5	9.5	238	4	US-09-495-880A-42	Sequence 42, Appl
44	73.5	9.5	266	4	US-09-495-880A-26	Sequence 26, Appl
45	73.5	9.5	293	3	US-08-438-745-4	Sequence 4, Appli

ALIGNMENTS

RESULT 1.  
US-08-233-788A-59  
; Sequence 59, Application US/08233788A  
; Patent No. 5635617  
; GENERAL INFORMATION:  
; APPLICANT: Doran, James L.  
; APPLICANT: Kay, William W.  
; APPLICANT: Collinson, Karen S.  
; APPLICANT: Clouthier, Sharon C.  
; TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR DETECTION  
; TITLE OF INVENTION: OF SALMONELLA  
; NUMBER OF SEQUENCES: 61  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Seed and Berry  
; STREET: 6300 Columbia Center, 701 Fifth Avenue  
; CITY: Seattle  
; STATE: Washington  
; COUNTRY: U.S.A.  
; ZIP: 98104-7092  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC Compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA: US/08/233,788A  
; FILING DATE: 26-APR-1994  
; CLASSIFICATION: 435  
; ATTORNEY/AGENT INFORMATION:  
; NAME: King, Joshua  
; REGISTRATION NUMBER: 35,570  
; REFERENCE/DOCKET NUMBER: 920043.403C2  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (206) 622-4900  
; TELEFAX: (206) 682-6031  
; TELEX: 3723836 SEEDANBERRY  
; INFORMATION FOR SEQ ID NO: 59:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 151 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
US-08-233-788A-59

Query Match 88.4%; Score 684; DB 1; Length 151;  
Best Local Similarity 90.1%; Pred. No. 5.5e-66;  
Matches 136; Conservative 3; Mismatches 12; Indels 0; Gaps 0;

QY 1 MKLLKVAFAAIVVSGSALAGVVPQGGGNGGNSGPDSTLSIYQYGSAHALAIQ 60

```
Db 1 MKLLKVAFAAIVVSGSALAGVVPQGGGNNHGGNSGPDSTLSIYQYGSANAALALQ 60
QY 61 SDARKYDQLVTRVVTTHMAHAGGADNSTLTQNGFRNNATIDQWNAKNSDITVGOYGG 120
Db 61 SDARKSETTITQSGYNGADVGCGADNSTLTQNGFRNNATIDQWNAKNSDITVGOYGG 120
QY 121 NNAALVNOTASDSSVMVRQVGFNNATANQY 151
Db 121 NNPALVNOTASDSSVMVRQVGFNNATANQY 151

RESULT 2
US-08-233-788A-57
; Sequence 57, Application US/08233788A
; Patent No. 5635617
; GENERAL INFORMATION:
; APPLICANT: Doran, James L.
; APPLICANT: Kay, William W.
; APPLICANT: Collinson, Karen S.
; APPLICANT: Clouthier, Sharon C.
; TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR DETECTION
; TITLE OF INVENTION: OF SALMONELLA
; NUMBER OF SEQUENCES: 61
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Seed and Berry
; STREET: 6300 Columbia Center, 701 Fifth Avenue
; CITY: Seattle
; STATE: Washington
; COUNTRY: U.S.A.
; ZIP: 98104-7092
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/233,788A
; FILING DATE: 26-APR-1994
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: King, Joshua
; REGISTRATION NUMBER: 35,570
; REFERENCE/DOCKET NUMBER: 920043.403C2
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (206) 622-4900
; TELEFAX: (206) 682-6031
; TELEX: 3723836 SEEDANBERRY
; INFORMATION FOR SEQ ID NO: 57:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 120 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-233-788A-57

Query Match 65.1%; Score 504; DB 1; Length 120;
Best Local Similarity 87.5%; Pred. No. 9.2e-47;
Matches 98; Conservative 3; Mismatches 11; Indels 0; Gaps 0;

QY 22 VVPQGGGNNHGGNSGPDSTLSIYQYGSANAALALQSDARKYDQLVTRVVTTHMAHA 81
Db 1 VVPQGGGNNHGGNSGPDSTLSIYQYGSANAALALQSDARKSETTITQSGYNGADV 60
QY 82 GQADNSTLTQNGFRNNATIDQWNAKNSDITVGOYGGNNAALVNOTASDS 133
Db 61 GQADNSTLTQNGFRNNATIDQWNAKNSDITVGOYGGNNAALVNOTASDS 112

RESULT 3
US-09-252-991A-27661
; Sequence 27661, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 27661
; LENGTH: 208
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-27661

Query Match 10.8%; Score 83.5; DB 4; Length 208;
Best Local Similarity 29.0%; Pred. No. 0.35;
Matches 31; Conservative 12; Mismatches 49; Indels 15; Gaps 3;

QY 3 LLKVAFAAIVVSGSAL-----AGVVPQGGGNNHGGNSGPDST-----LSIY 48
Db 101 LFAVAAL-GLLFAGSTLLPGELSPVGSAYAKGNGGNGGCHSGKGHGNLGGHSSK 159
QY 49 QYGSANAALALQSDARKYDQLVTRVVTTHMAHAGGADNSTLTION 95
Db 160 GHGSATSGIASRSDRGLSQASAIATTPGDHNSKGLSNAIGSTKN 206

RESULT 4
US-08-864-038A-3
; Sequence 3, Application US/08864038A
; Patent No. 6001592
; GENERAL INFORMATION:
; APPLICANT: Kunio NAKASHIMA et al.
; TITLE OF INVENTION: NOVEL POLYPEPTIDE GENE CDNA, VECTOR
; TITLE OF INVENTION: CONTAINING SAID CDNA, HOST CELLS TRANSFORMED WITH SAID
; TITLE OF INVENTION: VECTOR, POLYPEPTIDE PRODUCED THEREBY, METHOD OF PRODUCING
; TITLE OF INVENTION: SAID POLYPEPTIDE, DNA ENCODING SAID POLYPEPTIDE AND ANTIBODY
; TITLE OF INVENTION: TO SAID POLYPEPTIDE
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: 812-5 Hirano
; STREET: Isshinden
; CITY: Tsu-city
; STATE: Mie-prefecture
; COUNTRY: JAPAN
; ZIP: 514-01
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.50 inch, 1.44 MB storage
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: Microsoft Windows 95
; SOFTWARE: Word Perfect 6.1
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/864,038A
; FILING DATE: May 28, 1997
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: JP 8-184459
; FILING DATE: 15-July-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: C. Bruce Hamburg
; REGISTRATION NUMBER: 22,389
; REFERENCE/DOCKET NUMBER: F-5610
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212)986-2340
; TELEFAX: (212)953-7733
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 738
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
```



ORIGINAL SOURCE:  
 ORGANISM: Pinctada fucata  
 CELL TYPE: mantle epithelial cell  
 FEATURE:  
 NAME/KEY: peptide  
 LOCATION: from 1 to 738  
 IDENTIFICATION METHOD: E (by experiment)  
 US-08-864-038A-3

Query Match 10.8%; Score 83.5; DB 3; Length 738;  
 Best Local Similarity 27.7%; Pred. No. 2;  
 Matches 44; Conservative 12; Mismatches 56; Indels 47; Gaps 6;  
 QY 3 LKVAAPALIVVSGSALAGVWPQWGGGNGHNGSSGPDSTLSIYQYGSANAALALQSD 62  
 DB 419 LKSSASASASASAG-----GGGGGNGGGGGGG-----GCALAAALAA 465  
 QY 63 ARKYDQ-----VTRVVTHMAHAGGADNSTIELTQNGFRNNATIDOWNAKNSDITVQY 118  
 DB 466 AGAGGGLGGGGGALAAALAAAGAG-----GGFGGLGL-----GGL 504  
 QY 119 GGNNAALVNOTASDSS-----VMVQVGFNNATA 148  
 DB 505 GGGSAALAAAAAASGGGRALRRALRQMRGGGSA 543

## RESULT 5

US-09-498-520A-18  
 ; Sequence 18, Application US/09498520A  
 ; Patent No. 6613553  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Rock, Charles O  
 ; APPLICANT: Heath, Richard J  
 ; TITLE OF INVENTION: No. 6613553el Enoyl Reductases and Methods of Use Thereof  
 ; FILE REFERENCE: SJ-0022  
 ; CURRENT APPLICATION NUMBER: US/09/498, 520A  
 ; CURRENT FILING DATE: 2000-02-04  
 ; NUMBER OF SEQ ID NOS: 62  
 ; SOFTWARE: Patent in version 3.1  
 ; SEQ ID NO 18  
 ; LENGTH: 321  
 ; TYPE: PRT  
 ; ORGANISM: Caulobacter crescentus  
 US-09-498-520A-18

Query Match 10.3%; Score 79.5; DB 4; Length 321;  
 Best Local Similarity 24.5%; Pred. No. 1.7;  
 Matches 48; Conservative 15; Mismatches 64; Indels 69; Gaps 10;  
 QY 4 LKVAFAAIVVSGSA-----LAG-----VVPQWGGGNGHNG----- 34  
 DB 109 LKAAGLKVWVVCAGVAVKAEQAGCDAVICQGGGGGHTGLVGTLPVQAQVEAVKIPV 168  
 QY 35 ---GNSSSGPDSTLSIYQYGSANAALALQSDARKYDQVTRVVTHMAHAG-----Q 83  
 DB 169 VAAGGLHDG-----RGIA-LAALALGAGQ---VMVGTFTASHEAHAGDLYRQAVVE 215  
 QY 84 GADNSTIEL-TQNG-----FRNNATIDOWNAKNSDITV-----GYGGNNAALVN 127  
 DB 216 AADSDTVTRCYSGKPMRVKKNPVYDDWEARPGDIQPPQQAAMVSRNGAMGGIGGQIEG 275  
 QY 128 QTASDSSVMVQVCFG 143  
 DB 276 LDAAKSCFAMGQSAGG 291

## RESULT 6

US-09-328-352-4764  
 ; Sequence 4764, Application US/09328352  
 ; Patent No. 6562958  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Gary L. Breton et al.  
 ; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO ACINETOBACTER

; TITLE OF INVENTION: BAUMANNII FOR DIAGNOSTICS AND THERAPEUTICS.  
 ; FILE REFERENCE: GTC99-03PA  
 ; CURRENT APPLICATION NUMBER: US/09/328,352  
 ; CURRENT FILING DATE: 1999-06-04  
 ; NUMBER OF SEQ ID NOS: 8252  
 ; SEQ ID NO 4764  
 ; LENGTH: 975  
 ; TYPE: PRT  
 ; ORGANISM: Acinetobacter baumannii  
 US-09-328-352-4764

Query Match 10.3%; Score 79.5; DB 4; Length 975;  
 Best Local Similarity 25.0%; Pred. No. 8;  
 Matches 34; Conservative 19; Mismatches 70; Indels 13; Gaps 4;  
 QY 15 SGSALAGVWPQWGGGNGHNGSSGPDSTLSIYQYGSANAALALQSDARKYDQVTRVV 74  
 DB 279 NGTDSG-VSALGGSGNGSGDGAGNGIASNGEHNIGNG---NGDDVDITAPITGVL 333  
 QY 75 THE-----MAHAGGADNSTIELTQNGFRNNATIDOWNAKNSDITVQYGGNNAALVNQ 128  
 DB 334 NISGNSFTLIGNSSSSSVNTAPTTSNTVNDNTID--NGNSGGTSGSGNGSGDGLING 391  
 QY 129 TASDSSVMVQVCFG 144  
 DB 392 AASNGEHNIGNGN 407

## RESULT 7

US-08-458-023B-6  
 ; Sequence 6, Application US/08458023B  
 ; Patent No. 5667990  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Berka, Randy M.  
 ; APPLICANT: Yoder, Wendy  
 ; APPLICANT: Takagi, Shinobu  
 ; APPLICANT: Boominathan, Karuppan C.  
 ; TITLE OF INVENTION: ASPERGILLUS EXPRESSION SYSTEM  
 ; NUMBER OF SEQUENCES: 16  
 ; CORRESPONDENCE ADDRESS:  
 ; ADDRESSEE: No. 5667990o No. 5667990disk of No. 5667990th America, Inc.  
 ; STREET: 405 Lexington Avenue  
 ; CITY: New York  
 ; STATE: New York  
 ; COUNTRY: USA  
 ; ZIP: 10174-6201  
 ; COMPUTER READABLE FORM:  
 ; MEDIUM TYPE: Floppy disk  
 ; COMPUTER: IBM PC compatible  
 ; OPERATING SYSTEM: PC-DOS/MS-DOS  
 ; SOFTWARE: Patent In Release #1.0, Version #1.25  
 ; CURRENT APPLICATION DATA:  
 ; APPLICATION NUMBER: US/08/458,023B  
 ; FILING DATE: 01-JUN-1995  
 ; CLASSIFICATION: 435  
 ; ATTORNEY/AGENT INFORMATION:  
 ; NAME: Lowney Dr., Karen A.  
 ; REGISTRATION NUMBER: 31,274  
 ; REFERENCE/DOCKET NUMBER: 4086.010-US  
 ; TELECOMMUNICATION INFORMATION:  
 ; TELEPHONE: 212-867-0123  
 ; TELEFAX: 212-878-9655  
 ; INFORMATION FOR SEQ ID NO: 6:  
 ; SEQUENCE CHARACTERISTICS:  
 ; LENGTH: 363 amino acids  
 ; TYPE: amino acid  
 ; TOPOLOGY: linear  
 ; MOLECULE TYPE: protein  
 US-08-458-023B-6

Query Match 10.1%; Score 78; DB 1; Length 363;  
 Best Local Similarity 30.7%; Pred. No. 2.9;  
 Matches 35; Conservative 13; Mismatches 28; Indels 38; Gaps 8;

```

QY 1 MKLLKVAFAAIVVSGSALAGVVPQWGGGNH-----NGGNSSG-----PDSTLSI 47
   |||:||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 1 MKLSLLSTFAAVIIGALAL-----PQPGGGGSVTCFPGQSTNSQCCVWFVDLDDLQTNF 56
   |||:||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:

QY 48 YOYGSANAALALQSDARKYDQLVTRVWTHE-----MAHAGQ-----CADNSTI 90
   |||:||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 57 YQ-GS-----KCESPVK-----ILRIVFHDAlGFSPALTAAGQFGGGADGSII 100
   |||:||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:

RESULT 8
US-08-331-515A-2
; Sequence 2, Application US/08331515A
; Patent No. 5851811
; GENERAL INFORMATION:
; APPLICANT: Welinder, Karen
; APPLICANT: Andersen, Morten B
; TITLE OF INVENTION: PEROXIDASE VARIANTS WITH IMPROVED
; TITLE OF INVENTION: HYDROGEN PEROXIDE STABILITY
; NUMBER OF SEQUENCES: 2
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: No. 58518110 No. 585181ldisk of No. 585181lth America, Inc
; STREET: 405 Lexington Avenue, 64th Floor
; CITY: New York
; STATE: New York
; COUNTRY: United States of America
; ZIP: 10174-6401
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/331,515A
; FILING DATE: 01-NOV-1994
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Harrington, James J.
; REGISTRATION NUMBER: 38,711
; REFERENCE/DOCKET NUMBER: 3769.204-US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 212-867-0123
; TELEFAX: 212-878-9655
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 435 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
;
US-08-331-515A-2

Query Match 10.1%; Score 78; DB 2; Length 435;
Best Local Similarity 30.7%; Pred. No. 3.8;
Matches 35; Conservative 13; Mismatches 28; Indels 38; Gaps

QY 1 MKLLKVAFAAIVVSGSALAGVVPQWGGGNH-----NGGNSSG-----PDSTLSI 47
   |||:||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 2 MKLSLLSTFAAVIIGALAL-----PQPGGGGSVTCFPGQSTNSQCCVWFVDLDDLQTNF 57
   |||:||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:

QY 48 YOYGSANAALALQSDARKYDQLVTRVWTHE-----MAHAGQ-----CADNSTI 90
   |||:||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 58 YQ-GS-----KCESPVK-----ILRIVFHDAlGFSPALTAAGQFGGGADGSII 101
   |||:||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:

RESULT 9
US-09-168-406A-2
; Sequence 2, Application US/09168406A
; Patent No. 6258769
; GENERAL INFORMATION:
; APPLICANT: Welinder, Karen G.
; APPLICANT: Andersen, Morten B.
; TITLE OF INVENTION: Peroxidase Variants With Improved
; TITLE OF INVENTION: Hydrogen Peroxidase Stability

```

```
RESULT 11
US-09-823-494-26
; Sequence 26, Application US/09823494
; Patent No. 6355610
; GENERAL INFORMATION:
; APPLICANT: Chesebro, Bruce W
; APPLICANT: Caughey, Byron W
; APPLICANT: Chabry, Joelle
; APPLICANT: Priola, Susette
; TITLE OF INVENTION: Inhibitors of Formation of Protease Resistant Prion
; FILE REFERENCE: 50121
; CURRENT APPLICATION NUMBER: US/09/823,494
; PRIOR FILING DATE: 2001-03-30
; PRIOR APPLICATION NUMBER: 09/128,450
; PRIOR FILING DATE: 1998-08-03
; NUMBER OF SEQ ID NOS: 29
; SOFTWARE: Patent in Ver. 2.0
; SEQ ID NO 26
; LENGTH: 254
; TYPE: PRT
; ORGANISM: Hamster sp.
US-09-823-494-26

Query Match      10.0%; Score 77.5; DB 4; Length 254;
Best Local Similarity 25.4%; Pred. No. 2;
Matches 32; Conservative 19; Mismatches 60; Indels 15; Gaps 6;

QY 26 WG-GGNHNGGNSGPDSTLSTIYQYGSANAALQSDARKYDQLVTRVVTHMAHAGOG 84
DB 89 WGGGGTHQWKNKPSKPNMK--HWAGAAAGAVVGGGLGY--MLGSAMSPMKHFGND 144
QY 85 ADNSTIELTQNGFRNNA---TIDWNAKNS-----DITVGYGGNNAAL-VNQASDSS 134
DB 145 WEDRYVRENWNRYPNQVYRPVDQYNNQNNVHDCVNIKQHTVTTTKGENFTETDIK 204
QY 135 VMVRQV 140
DB 205 IMERVV 210

RESULT 12
US-07-792-259-17
; Sequence 17, Application US/07792259
; Patent No. 5286638
; GENERAL INFORMATION:
; APPLICANT: TANAKA, YOSHIKAZU
; APPLICANT: ASHIKARI, TOSHIHIKO
; APPLICANT: HATANAKA, HARUYO
; APPLICANT: SHIBANO, YUJI
; APPLICANT: AMACHI, TERUO
; APPLICANT: NAKAYAMA, TORU
; APPLICANT: SUMIDA, MOTOO
; TITLE OF INVENTION: PEROXIDE GENE OF MICROBIAL ORIGIN
; NUMBER OF SEQUENCES: 17
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: CUSHMAN, DABBY & CUSHMAN
; STREET: 1615 L. STREET, N.W.
; CITY: WASHINGTON
; STATE: D.C.
; COUNTRY: USA
; ZIP: 20036
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent in Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/792,259
; FILING DATE: 19911115
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: SCOTT, WATSON T.
```

```
; REGISTRATION NUMBER: 26581
; REFERENCE/DOCKET NUMBER: 9437/93433
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 202-861-3067
; TELEFAX: 202-822-0944
; TELEX: 6714627 CUSH
; INFORMATION FOR SEQ ID NO: 17:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 364 amino acids
; TYPE: AMINO ACID
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
US-07-792-259-17

Query Match      10.0%; Score 77.5; DB 1; Length 364;
Best Local Similarity 30.4%; Pred. No. 3.3;
Matches 35; Conservative 13; Mismatches 28; Indels 39; Gaps 8;

QY 1 MLLKVAAPAAIVVSGSALAGVVPWGCGGNHN---GGGNSSG-----PDSTLS 46
DB 1 MKLSLFSTFAAVIIGALAL----PQGFGGGGSVTCFGGOSTNSQCCVWFVDVLDLQTN 56
QY 47 IYQYGSANAALQSDARKYDQLVTRVVTHMAHAGQ---GADNSTI 90
DB 57 FYQ-GS-----XCSPVRK----ILRVFHDAGFSPALTAAGQFGGGGADGSI 101

RESULT 13
US-09-056-556-204
; Sequence 204, Application US/09056556
; Patent No. 6350456
; GENERAL INFORMATION:
; APPLICANT: Reed, Steven G.
; APPLICANT: Skeiky, Yasir A.W.
; APPLICANT: Dillon, Davin C.
; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR THE PREVENTION AND
; NUMBER OF SEQUENCES: 241
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SEED and BERRY LLP
; STREET: 6300 Columbia Center, 701 Fifth Avenue
; CITY: Seattle
; STATE: Washington
; COUNTRY: USA
; ZIP: 98104-7092
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent in Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/056,556
; FILING DATE: 07-APR-1998
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Maki, David J.
; REGISTRATION NUMBER: 31,392
; REFERENCE/DOCKET NUMBER: 210121.457
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (206) 622-4900
; TELEFAX: (206) 682-6031
; INFORMATION FOR SEQ ID NO: 204:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 943 amino acids
; TYPE: amino acid
; STRANDEDNESS:
; TOPOLOGY: linear
US-09-056-556-204

Query Match      9.9%; Score 77; DB 4; Length 943;
Best Local Similarity 26.2%; Pred. No. 14;
Matches 37; Conservative 14; Mismatches 66; Indels 24; Gaps 7;
```

TREATME

```

QY 16 GSAAGVVPQWGGG-GNHN-GGNSGSPDSTLSIYQYGSANAA--LALQSDARKYDQLVT 71
Db 464 GSGNIGVFNVGSGSLGNYNIGSGN-----LGIYNIGFNGVDYVNGFNGAGDFNQGFA 516
QY 72 RVVTHEMAHAGQAGADNSTIELT---QNGFRNNATIDQWNAKNSDITVQYGGNNAALVNO 128
Db 517 NTGNNNIGFANTGNNNIGLSGDNQOQFN---IASGWSGTGNSGLFNSGTNNVGI FNA 573
QY 129 TASDSSVMVRQVGFGNANAT 149
Db 574 GTGN-----VGIANS GTGN 587

RESULT 14
US-09-072-596-199
; Sequence 199, Application US/09072596
; Patent No. 6458366
; GENERAL INFORMATION:
; APPLICANT: Reed, Steven G.
; APPLICANT: Skeiky, Yasir A.W.
; APPLICANT: Dillon, Davin C.
; APPLICANT: Campos-Neto, Antonia
; APPLICANT: Houghton, Raymond
; APPLICANT: Vedvick, Thomas S.
; APPLICANT: Twardzik, Daniel R.
; APPLICANT: Lodes, Michael J.
; APPLICANT: Hendrickson, Ronald C.
; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR DIAGNOSIS OF TUBERCULOSIS
; NUMBER OF SEQUENCES: 350
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SEED and BERRY LLP
; STREET: 6300 Columbia Center, 701 Fifth Avenue
; CITY: Seattle
; STATE: Washington
; COUNTRY: USA
; ZIP: 98104-7092
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent in Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/072,596
; FILING DATE: 05-MAY-1998
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Maki, David J.
; REGISTRATION NUMBER: 31,392
; REFERENCE/DOCKET NUMBER: 210121.417C9
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (206) 622-4900
; TELEFAX: (206) 682-6031
; INFORMATION FOR SEQ ID NO: 199:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 943 amino acids
; TYPE: amino acid
; STRANDEDNESS:
; TOPOLOGY: linear
US-09-072-596-199

```

```

Query Match 9.9%; Score 77; DB 4; Length 943;
Best Local Similarity 26.2%; Pred. No. 14;
Matches 37; Conservative 14; Mismatches 66; Indels 24; Gaps 7;

QY 16 GSAAGVVPQWGGG-GNHN-GGNSGSPDSTLSIYQYGSANAA--LALQSDARKYDQLVT 71
Db 464 GSGNIGVFNVGSGSLGNYNIGSGN-----LGIYNIGFNGVDYVNGFNGAGDFNQGFA 516
QY 72 RVVTHEMAHAGQAGADNSTIELT---QNGFRNNATIDQWNAKNSDITVQYGGNNAALVNO 128
Db 517 NTGNNNIGFANTGNNNIGLSGDNQOQFN---IASGWSGTGNSGLFNSGTNNVGI FNA 573
QY 129 TASDSSVMVRQVGFGNANAT 149

```

```

Db 574 GTGN-----VGIANS GTGN 587

RESULT 15
US-09-477-135A-131
; Sequence 131, Application US/09477135A
; Patent No. 6572865
; GENERAL INFORMATION:
; APPLICANT: Naro, Francis
; TITLE OF INVENTION: Mycobacterium Tuberculosis DNA Sequences Encoding
; TITLE OF INVENTION: Immunostimulatory Peptides
; FILE REFERENCE: 52888
; CURRENT APPLICATION NUMBER: US/09/477,135A
; CURRENT FILING DATE: 2000-01-03
; PRIOR APPLICATION NUMBER: 08990823
; PRIOR FILING DATE: 1997-12-15
; PRIOR APPLICATION NUMBER: US 96/10375
; PRIOR FILING DATE: 1996-06-14
; PRIOR APPLICATION NUMBER: 60/000,254
; PRIOR FILING DATE: 1995-06-15
; NUMBER OF SEQ ID NOS: 169
; SOFTWARE: Patent in Ver. 2.0
; SEQ ID NO 131
; LENGTH: 943
; TYPE: PRT
; ORGANISM: Mycobacterium tuberculosis
US-09-477-135A-131

Query Match 9.9%; Score 77; DB 4; Length 943;
Best Local Similarity 26.2%; Pred. No. 14;
Matches 37; Conservative 14; Mismatches 66; Indels 24; Gaps 7;

QY 16 GSAAGVVPQWGGG-GNHN-GGNSGSPDSTLSIYQYGSANAA--LALQSDARKYDQLVT 71
Db 571 GSGNIGVFNVGSGSLGNYNIGSGN-----LGIYNIGFNGVDYVNGFNGAGDFNQGFA 623
QY 72 RVVTHEMAHAGQAGADNSTIELT---QNGFRNNATIDQWNAKNSDITVQYGGNNAALVNO 128
Db 624 NTGNNNIGFANTGNNNIGLSGDNQOQFN---IASGWSGTGNSGLFNSGTNNVGI FNA 680
QY 129 TASDSSVMVRQVGFGNANAT 149
Db 681 GTGN-----VGIANS GTGN 694

Search completed: August 2, 2004, 14:58:34
Job time : 13 secs

```

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: August 2, 2004, 14:54:48 ; Search time 36.8 Seconds  
(without alignments)  
1287.123 Million cell updates/sec

Title: US-09-543-407-20

Perfect score: 774

Sequence: 1 MKLLKVAFAAIVVSGSALA.....DSSVMVRQVGFNNATANYQ 151

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1291235 seqs, 313682936 residues

Total number of hits satisfying chosen parameters: 1291235

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published Applications AA:\*

- 1: /cgn2\_6/ptodata/2/pubpaa/US07\_PUBCOMB.pep:\*
- 2: /cgn2\_6/ptodata/2/pubpaa/PCT\_NEW\_PUB.pep:\*
- 3: /cgn2\_6/ptodata/2/pubpaa/US06\_NEW\_PUB.pep:\*
- 4: /cgn2\_6/ptodata/2/pubpaa/US06\_PUBCOMB.pep:\*
- 5: /cgn2\_6/ptodata/2/pubpaa/US07\_NEW\_PUB.pep:\*
- 6: /cgn2\_6/ptodata/2/pubpaa/PCTUS\_PUBCOMB.pep:\*
- 7: /cgn2\_6/ptodata/2/pubpaa/US08\_NEW\_PUB.pep:\*
- 8: /cgn2\_6/ptodata/2/pubpaa/US08\_PUBCOMB.pep:\*
- 9: /cgn2\_6/ptodata/2/pubpaa/US09A\_PUBCOMB.pep:\*
- 10: /cgn2\_6/ptodata/2/pubpaa/US09B\_PUBCOMB.pep:\*
- 11: /cgn2\_6/ptodata/2/pubpaa/US09C\_PUBCOMB.pep:\*
- 12: /cgn2\_6/ptodata/2/pubpaa/US09\_NEW\_PUB.pep:\*
- 13: /cgn2\_6/ptodata/2/pubpaa/US10A\_PUBCOMB.pep:\*
- 14: /cgn2\_6/ptodata/2/pubpaa/US10B\_PUBCOMB.pep:\*
- 15: /cgn2\_6/ptodata/2/pubpaa/US10C\_PUBCOMB.pep:\*
- 16: /cgn2\_6/ptodata/2/pubpaa/US10\_NEW\_PUB.pep:\*
- 17: /cgn2\_6/ptodata/2/pubpaa/US60\_NEW\_PUB.pep:\*
- 18: /cgn2\_6/ptodata/2/pubpaa/US60\_PUBCOMB.pep:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	525	67.8	151	12 US-09-741-873B-4	Sequence 4, Appli
2	525	67.8	151	12 US-09-741-873B-4	Sequence 4, Appli
3	447	57.8	131	12 US-09-741-873B-2	Sequence 2, Appli
4	447	57.8	131	12 US-09-741-873B-2	Sequence 2, Appli
5	101.5	13.1	445	15 US-10-369-493-20638	Sequence 20638, A
6	94.5	12.2	438	14 US-10-156-761-9343	Sequence 9343, Ap
7	91	11.8	186	12 US-10-282-122A-49412	Sequence 49412, A
8	85	11.0	3705	12 US-10-282-122A-77944	Sequence 77944, A
9	85	11.0	6310	12 US-10-282-122A-67793	Sequence 67793, A
10	84.5	10.9	1778	14 US-10-238-075-749	Sequence 749, App
11	84	10.9	1129	12 US-10-282-122A-48048	Sequence 48048, A
12	83.5	10.8	278	9 US-09-810-264-28	Sequence 28, Appl
13	83	10.7	597	9 US-09-793-306-146	Sequence 146, App
14	83	10.7	678	12 US-10-282-122A-64573	Sequence 64573, A
15	82.5	10.7	435	14 US-10-128-714-3213	Sequence 3213, Ap

16	82.5	10.7	515	14 US-10-128-714-8213	Sequence 8213, Ap
17	82.5	10.7	2732	14 US-10-238-075-1119	Sequence 1119, Ap
18	82.5	10.7	2834	14 US-10-085-959-252	Sequence 252, App
19	82	10.6	85	9 US-09-996-194-16	Sequence 16, Appl
20	82	10.6	85	12 US-10-164-965-33	Sequence 33, Appl
21	82	10.6	354	10 US-09-820-843A-21	Sequence 21, Appl
22	82	10.6	440	12 US-10-424-599-229222	Sequence 229222, A
23	82	10.6	1862	12 US-10-282-122A-49757	Sequence 49757, A
24	81.5	10.5	2204	12 US-10-282-122A-64364	Sequence 64364, A
25	81	10.5	562	12 US-10-282-122A-64514	Sequence 64514, A
26	81	10.5	628	12 US-10-282-122A-53269	Sequence 53269, A
27	80.5	10.4	154	16 US-10-437-963-182284	Sequence 162284, A
28	80.5	10.4	486	15 US-10-369-493-20619	Sequence 20619, A
29	80	10.3	145	16 US-10-437-963-147748	Sequence 147748, A
30	80	10.3	477	12 US-10-425-114-70098	Sequence 70098, A
31	80	10.3	507	12 US-10-424-599-229226	Sequence 229226, A
32	80	10.3	538	12 US-10-425-114-68152	Sequence 68152, A
33	80	10.3	558	16 US-10-437-963-175203	Sequence 175203, A
34	79.5	10.3	562	14 US-10-156-761-13039	Sequence 13039, A
35	79.5	10.3	1721	12 US-10-282-122A-62548	Sequence 62548, A
36	79	10.2	209	12 US-10-424-599-221110	Sequence 221110, A
37	79	10.2	276	15 US-10-369-493-3641	Sequence 3641, Ap
38	79	10.2	688	14 US-10-032-585-7876	Sequence 7876, Ap
39	79	10.2	974	12 US-10-282-122A-44999	Sequence 44999, A
40	78.5	10.1	271	14 US-10-156-761-11721	Sequence 11721, A
41	78.5	10.1	292	16 US-10-437-963-195404	Sequence 195404, A
42	78.5	10.1	472	16 US-10-467-479-2	Sequence 2, Appli
43	78	10.1	545	15 US-10-369-493-18473	Sequence 18473, A
44	78	10.1	594	14 US-10-156-761-13173	Sequence 13173, A
45	78	10.1	1246	12 US-10-282-122A-49773	Sequence 49773, A

#### ALIGNMENTS

#### RESULT 1

US-09-741-873B-4  
; Sequence 4, Application US/09741873B  
; Publication No. US20020081722A1  
; GENERAL INFORMATION:  
; APPLICANT: Normark, Staffan  
; APPLICANT: Olsen, Arne  
; FILE OF INVENTION: Fibronectin Binding Protein As Well As Its Preparation  
; TITLE REFERENCE: 012889-084  
; CURRENT APPLICATION NUMBER: US/09/741,873B  
; CURRENT FILING DATE: 2003-04-04  
; PRIOR FILING DATE: 1998-05-06  
; PRIOR APPLICATION NUMBER: SE 8801723-1  
; PRIOR FILING DATE: 1997-11-26  
; PRIOR APPLICATION NUMBER: US 08/978,878  
; PRIOR FILING DATE: 1997-11-26  
; PRIOR APPLICATION NUMBER: US 07/347,189  
; PRIOR FILING DATE: 1989-05-04  
; PRIOR APPLICATION NUMBER: US 07/789,437  
; PRIOR FILING DATE: 1991-11-06  
; PRIOR APPLICATION NUMBER: US 07/970,846  
; PRIOR FILING DATE: 1992-11-03  
; PRIOR APPLICATION NUMBER: US 08/187,865  
; PRIOR FILING DATE: 1994-01-28  
; PRIOR APPLICATION NUMBER: US 08/318,519  
; PRIOR FILING DATE: 1994-10-05  
; NUMBER OF SEQ ID NOS: 10  
; SOFTWARE: Patent in version 3.0  
; SEQ ID NO 4  
; LENGTH: 151  
; TYPE: PRT  
; ORGANISM: Escherichia coli  
US-09-741-873B-4

Query Match 67.8%; Score 525; DB 12; Length 151;  
Best Local Similarity 68.2%; Pred. No. 6e-46;  
Matches 103; Conservative 21; Mismatches 27; Indels 0; Gaps 0;

QY 1 MKLLKVAFAAIVVSGSALAGVWPQGGGNGHNGSSGPDSTLSIYQGSANAALALQ 60

```

Db      1  MKLLKVAALAAIVFSGSAGVVPQVGGGNGHGGGNSGPNSELNIYQYGGNSALALQ 60
QY      61  SDARKYDQLVTRVVTTHMAHAGGADNSTIETQNGFRNNAIDQWNAKNSDITVQYGG 120
Db      61  TDARNSDLITQHGNGGADVGQSDSSIDLITQRFNGSATLDQWNGKNSMTVKQFGG 120
QY      121  NNAALVNOTASDSSVMVRQVGFNNATANOY 151
Db      121  GNGAAVDQTASNSVNVTVQVGFNNATANOY 151

```

```

RESULT 2
US-09-741-873B-4
; Sequence 4, Application US/09741873B
; Publication No. US20040096965A9
; GENERAL INFORMATION:
; APPLICANT: Olsen, Stafan
; APPLICANT: Olsen, Arne
; TITLE OF INVENTION: Fibronectin Binding Protein As Well As Its Preparation
; FILE REFERENCE: 012889-084
; CURRENT APPLICATION NUMBER: US/09/741,873B
; CURRENT FILING DATE: 2003-04-04
; PRIOR APPLICATION NUMBER: SE 8801723-1
; PRIOR FILING DATE: 1998-05-06
; PRIOR APPLICATION NUMBER: US 08/978,878
; PRIOR FILING DATE: 1997-11-26
; PRIOR APPLICATION NUMBER: US 07/347,189
; PRIOR FILING DATE: 1989-05-04
; PRIOR APPLICATION NUMBER: US 07/789,437
; PRIOR FILING DATE: 1991-11-06
; PRIOR APPLICATION NUMBER: US 07/970,846
; PRIOR FILING DATE: 1992-11-03
; PRIOR APPLICATION NUMBER: US 08/187,865
; PRIOR FILING DATE: 1994-01-28
; PRIOR APPLICATION NUMBER: US 08/318,519
; NUMBER OF SEQ ID NOS: 10
; SOFTWARE: Patentin version 3.0
; SEQ ID NO 4
; LENGTH: 151
; TYPE: PRT
; ORGANISM: Escherichia coli
US-09-741-873B-4

```

```

Query Match      67.8%; Score 525; DB 12; Length 151;
Best Local Similarity 68.2%; Pred. No. 6e-46;
Matches 103; Conservative 21; Mismatches 27; Indels 0; Gaps 0;

QY      1  MKLLKVAALAAIVFSGSAGVVPQVGGGNGHGGGNSGPNSELNIYQYGGNSALALQ 60
Db      1  MKLLKVAALAAIVFSGSAGVVPQVGGGNGHGGGNSGPNSELNIYQYGGNSALALQ 60
QY      61  SDARKYDQLVTRVVTTHMAHAGGADNSTIETQNGFRNNAIDQWNAKNSDITVQYGG 120
Db      61  TDARNSDLITQHGNGGADVGQSDSSIDLITQRFNGSATLDQWNGKNSMTVKQFGG 120
QY      121  NNAALVNOTASDSSVMVRQVGFNNATANOY 151
Db      121  GNGAAVDQTASNSVNVTVQVGFNNATANOY 151

```

```

RESULT 3
US-09-741-873B-2
; Sequence 2, Application US/09741873B
; Publication No. US20020081722A1
; GENERAL INFORMATION:
; APPLICANT: Olsen, Stafan
; APPLICANT: Olsen, Arne
; TITLE OF INVENTION: Fibronectin Binding Protein As Well As Its Preparation
; FILE REFERENCE: 012889-084
; CURRENT APPLICATION NUMBER: US/09/741,873B
; CURRENT FILING DATE: 2003-04-04
; SEQ ID NO 4
; LENGTH: 151
; TYPE: PRT
; ORGANISM: Escherichia coli
US-09-741-873B-2

```

```

Query Match      67.8%; Score 525; DB 12; Length 151;
Best Local Similarity 68.2%; Pred. No. 6e-46;
Matches 103; Conservative 21; Mismatches 27; Indels 0; Gaps 0;

QY      1  MKLLKVAALAAIVFSGSAGVVPQVGGGNGHGGGNSGPNSELNIYQYGGNSALALQ 60
Db      1  MKLLKVAALAAIVFSGSAGVVPQVGGGNGHGGGNSGPNSELNIYQYGGNSALALQ 60
QY      61  SDARKYDQLVTRVVTTHMAHAGGADNSTIETQNGFRNNAIDQWNAKNSDITVQYGG 120
Db      61  TDARNSDLITQHGNGGADVGQSDSSIDLITQRFNGSATLDQWNGKNSMTVKQFGG 120
QY      121  NNAALVNOTASDSSVMVRQVGFNNATANOY 151
Db      121  GNGAAVDQTASNSVNVTVQVGFNNATANOY 151

```

```

; PRIOR APPLICATION NUMBER: SE 8801723-1
; PRIOR FILING DATE: 1998-05-06
; PRIOR APPLICATION NUMBER: US 08/978,878
; PRIOR FILING DATE: 1997-11-26
; PRIOR APPLICATION NUMBER: US 07/347,189
; PRIOR FILING DATE: 1989-05-04
; PRIOR APPLICATION NUMBER: US 07/789,437
; PRIOR FILING DATE: 1991-11-06
; PRIOR APPLICATION NUMBER: US 07/970,846
; PRIOR FILING DATE: 1992-11-03
; PRIOR APPLICATION NUMBER: US 08/187,865
; PRIOR FILING DATE: 1994-01-28
; PRIOR APPLICATION NUMBER: US 08/318,519
; NUMBER OF SEQ ID NOS: 10
; SOFTWARE: Patentin version 3.0
; SEQ ID NO 2
; LENGTH: 131
; TYPE: PRT
; ORGANISM: Escherichia coli
US-09-741-873B-2

```

```

Query Match      57.8%; Score 447; DB 12; Length 131;
Best Local Similarity 64.9%; Pred. No. 5.2e-38;
Matches 85; Conservative 20; Mismatches 26; Indels 0; Gaps 0;

QY      21  GVVPQWGGGNGHGGGNSGPNSELNIYQYGGNSALALQSDARKYDQLVTRVVTTHMAH 80
Db      1  GVVPQWGGGNGHGGGNSGPNSELNIYQYGGNSALALQSDARKYDQLVTRVVTTHMAH 80
QY      81  AGGADNSTIETQNGFRNNAIDQWNAKNSDITVQYGGNSALALQSDARKYDQLVTRVVTTHMAH 140
Db      61  VGGSDSSIDLITQRFNGSATLDQWNGKNSMTVKQFGGNGAAYDQTASNSVNVTVQ 120
QY      141  GFGNNATANOY 151
Db      121  GFGNNATANOY 131

```

```

RESULT 4
US-09-741-873B-2
; Sequence 2, Application US/09741873B
; Publication No. US20040096965A9
; GENERAL INFORMATION:
; APPLICANT: Olsen, Stafan
; APPLICANT: Olsen, Arne
; TITLE OF INVENTION: Fibronectin Binding Protein As Well As Its Preparation
; FILE REFERENCE: 012889-084
; CURRENT APPLICATION NUMBER: US/09/741,873B
; CURRENT FILING DATE: 2003-04-04
; PRIOR APPLICATION NUMBER: SE 8801723-1
; PRIOR FILING DATE: 1998-05-06
; PRIOR APPLICATION NUMBER: US 08/978,878
; PRIOR FILING DATE: 1997-11-26
; PRIOR APPLICATION NUMBER: US 07/347,189
; PRIOR FILING DATE: 1989-05-04
; PRIOR APPLICATION NUMBER: US 07/789,437
; PRIOR FILING DATE: 1991-11-06
; PRIOR APPLICATION NUMBER: US 07/970,846
; PRIOR FILING DATE: 1992-11-03
; PRIOR APPLICATION NUMBER: US 08/187,865
; PRIOR FILING DATE: 1994-01-28
; PRIOR APPLICATION NUMBER: US 08/318,519
; NUMBER OF SEQ ID NOS: 10
; SOFTWARE: Patentin version 3.0
; SEQ ID NO 2
; LENGTH: 131
; TYPE: PRT
; ORGANISM: Escherichia coli
US-09-741-873B-2

```

```

Query Match      57.8%; Score 447; DB 12; Length 131;

```

Best Local Similarity 64.9%; Pred. No. 5.2e-38;  
Matches 85; Conservative 20; Mismatches 26; Indels 0; Gaps 0;  
QY 21 GVVPQGGGNGGNSGSPSTLSIYQYGSANAALALOSDARKYDQLVTVVTHEMAH 80  
Db 1 GVVPQGGGNGGNGGNSGSPSTLSIYQYGSANAALALQTDARNSDLTITOHGGNGAD 60  
QY 81 AGGADNSTIETQNGFRNATIDWNKNSDITVQYGGNNAALVNQTASDSSVMVRQV 140  
Db 61 VGGGSDSSIDLTFQGFNSATLDQWNGKNSMTVKQFGGGNGAAVDQTASNSSVNVTVQ 120  
QY 141 GFGNNTANQY 151  
Db 121 GFGNNTANQY 131

RESULT 5  
US-10-369-493-20638  
; Sequence 20638, Application US/10369493  
; Publication No. US20030233675A1  
; GENERAL INFORMATION:  
; APPLICANT: Cao, Yongwei  
; APPLICANT: Hinkle, Gregory J.  
; APPLICANT: Slater, Steven C.  
; APPLICANT: Goldman, Barry S.  
; APPLICANT: Chen, Xianfeng  
; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF  
; FILE REFERENCE: 38-10(52052)B  
; CURRENT APPLICATION NUMBER: US/10/369,493  
; CURRENT FILING DATE: 2003-02-28  
; PRIOR APPLICATION NUMBER: US 60/360,039  
; PRIOR FILING DATE: 2002-02-21  
; NUMBER OF SEQ ID NOS: 47374  
; SEQ ID NO 20638  
; LENGTH: 445  
; TYPE: PRT  
; ORGANISM: Rhodospseudomonas palustris  
; FEATURE:  
; NAME/KEY: unsure  
; LOCATION: (1)..(445)  
; OTHER INFORMATION: unsure at all Xaa locations  
US-10-369-493-20638

Query Match 13.1%; Score 101.5; DB 15; Length 445;  
Best Local Similarity 26.0%; Pred. No. 0.076;  
Matches 45; Conservative 20; Mismatches 55; Indels 53; Gaps 7;  
QY 7 AAFAA-----IVVSGSALAGVVPQGGGG-----NHNGG-----GNSGPDSTLSIYQY 50  
Db 19 AAFADSNVTYVLTQNDQANITQSGNGSVGAFNGSGFLQNGTLGA-NLLTVKQS 77  
QY 51 GSANAALALOSDARKYDQLVTVVTHEMAHAGOGA-----DNSTIELTQNGFRNATID 104  
Db 78 GNSNS-----VGRDIQKQSGAGNSAIFQBTGSDVLEQOTGTSGNAVPS 123  
QY 105 QWNAKN-----SDITVQYGGNNAALVNQTASDSSVMVRQV 141  
Db 124 GAWTNDPGVFNKTDQSSNGSKSVIQQGKNVFSIKQGTGNTSTSVNQIG 176

RESULT 6  
US-10-156-761-9343  
; Sequence 9343, Application US/10156761  
; Publication No. US20030119018A1  
; GENERAL INFORMATION:  
; APPLICANT: OMURA, SATOSHI  
; APPLICANT: IKEDA, HARUO  
; APPLICANT: HORIKAWA, JUN  
; APPLICANT: HORIKAWA, HIROSHI  
; APPLICANT: SHIBA, TADAYOSHI  
; APPLICANT: SAKAKI, YOSHIYUKI  
; APPLICANT: HATTORI, MASAHIRA

; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES  
; FILE REFERENCE: 249-262  
; CURRENT APPLICATION NUMBER: US/10/156,761  
; CURRENT FILING DATE: 2002-05-29  
; PRIOR APPLICATION NUMBER: JP 2001-204089  
; PRIOR FILING DATE: 2001-05-30  
; PRIOR APPLICATION NUMBER: JP 2001-272697  
; PRIOR FILING DATE: 2001-08-02  
; NUMBER OF SEQ ID NOS: 15109  
; SEQ ID NO 9343  
; LENGTH: 438  
; TYPE: PRT  
; ORGANISM: Streptomyces avermitilis  
US-10-156-761-9343

Query Match 12.2%; Score 94.5; DB 14; Length 438;  
Best Local Similarity 25.9%; Pred. No. 0.39;  
Matches 37; Conservative 27; Mismatches 46; Indels 33; Gaps 8;  
QY 1 MKLLKVAFAAIVVSGSALAGVVPQGGGNGGNSGSPDSTLSIYQYGSANAALA-- 58  
Db 1 MRSIRAAAVGAVTWSLALASAC-----GGGSGTGGSDSP-KTLT-YWASNQASIAVD 54  
QY 59 ---LQSDARKYDQ-----LVTVVTHEMAHAGOGADNSTIELTQNGFRN 99  
Db 55 KKVLPQLDKFEETGKIVKLVVVPWSDLLNRLT--ATTSGQGPVNLNIGNTWSASLQA 112  
QY 100 NATIDWNKNSDITVQYGGN 122  
Db 113 TGALLPMDAKNFD----KIGKGD 131

RESULT 7  
US-10-282-122A-49412  
; Sequence 49412, Application US/10282122A  
; Publication No. US20040029129A1  
; GENERAL INFORMATION:  
; APPLICANT: Wang, Liangsu  
; APPLICANT: Zamudio, Carlos  
; APPLICANT: Malone, Cheryl  
; APPLICANT: Haselbeck, Robert  
; APPLICANT: Chisen, Karl  
; APPLICANT: Zyskind, Judith  
; APPLICANT: Wall, Daniel  
; APPLICANT: Trawick, John  
; APPLICANT: Carr, Grant  
; APPLICANT: Yamamoto, Robert  
; APPLICANT: Forsyth, R.  
; APPLICANT: Xu, H.  
; TITLE OF INVENTION: Identification of Essential Genes in Microorganisms  
; FILE REFERENCE: ELITRA.034A  
; CURRENT APPLICATION NUMBER: US/10/282,122A  
; CURRENT FILING DATE: 2003-02-20  
; PRIOR APPLICATION NUMBER: 60/191,078  
; PRIOR FILING DATE: 2000-03-21  
; PRIOR APPLICATION NUMBER: 60/206,848  
; PRIOR FILING DATE: 2000-05-23  
; PRIOR APPLICATION NUMBER: 60/207,727  
; PRIOR FILING DATE: 2000-05-26  
; PRIOR APPLICATION NUMBER: 60/230,335  
; PRIOR FILING DATE: 2000-09-06  
; PRIOR APPLICATION NUMBER: 60/230,347  
; PRIOR FILING DATE: 2000-09-09  
; PRIOR APPLICATION NUMBER: 60/242,578  
; PRIOR FILING DATE: 2000-10-23  
; PRIOR APPLICATION NUMBER: 60/253,625  
; PRIOR FILING DATE: 2000-11-27  
; PRIOR APPLICATION NUMBER: 60/257,931  
; PRIOR FILING DATE: 2000-12-22  
; PRIOR APPLICATION NUMBER: 60/267,636  
; PRIOR FILING DATE: 2001-03-09  
; PRIOR APPLICATION NUMBER: 60/269,308  
; PRIOR FILING DATE: 2001-02-16





```
QY 10 RAIVVSGSALAGVPOW-----GGGG-----NHNGGNSG---PDSILSIYQGSAN 54
Db 111 AGIGTAGAM-----PTWAIAGSLLGVGAAADNGGGSGSGSDPTS-----156
QY 55 AALALQSDARKYDQLVTRVVTHEMAHAGQADNSTIEL-----TONGFRNN 100
Db 157 -----APATPIDLLVS-----PDGLRLTGRGEAGTTVNIIRDAAGNLIGSGTVGADGNFNT 207
QY 101 ATIDQWNAKSDITVQYGGNNAALVNQTASDSSVMVQVGFNNATAN 149
Db 208 LNAPOINSENLDVTLTDAAGNVSAFCAVTAPDAPLAPTDLAINEQGN 256

RESULT 10
US-10-238-075-749
; Sequence 749, Application US/10238075
; Publication No. US20030149324A1
; GENERAL INFORMATION:
; APPLICANT: I.N.S.E.R.M.
; TITLE OF INVENTION: Polynucleotides which are of nature B2/D+ A- and which are isolated
; TITLE OF INVENTION: E.coli, and biological uses of these polynucleotides and of their
; FILE REFERENCE: BLANDINE
; CURRENT APPLICATION NUMBER: US/10/238,075
; CURRENT FILING DATE: 2002-09-10
; PRIOR APPLICATION NUMBER: 0003145
; PRIOR FILING DATE: 2000-03-10
; NUMBER OF SEQ ID NOS: 1576
; SOFTWARE: Patentin version 3.1
; SEQ ID NO 749
; LENGTH: 1778
; TYPE: PRT
; ORGANISM: Escherichia coli
US-10-238-075-749

Query Match 10.9%; Score 84.5; DB 14; Length 1778;
Best Local Similarity 26.8%; Pred. No. 25;
Matches 40; Conservative 21; Mismatches 59; Indels 29; Gaps 7;

QY 5 KVAFAAIVVSGSALAGVPOW-----GGGNHNGGNSGSDPTSILSIYQGSANALAL- 59
Db 906 KISSNSTDAINGSLYGVADSFSTYLGCGGADISDTGLSGPTTIGGDTYVNGDALAAI 965
QY 60 -----QSDARKYDQLVTRVVTHEMAHAGQADNSTIELTQ--NGFRNATIDQWNAKN 110
Db 966 NTSFSTSLGDLALWD-----ATAGKFSKXHGNNAPSVTIDVANGAVSSTSSDAINGSQ 1019

QY 111 ----SDITVQYGGNNAALVNQTASDSSV 135
Db 1020 LYGVSDYIADALGN--AVVN---TDGSI 1043

RESULT 11
US-10-282-122A-48048
; Sequence 48048, Application US/10282122A
; Publication No. US20040029129A1
; GENERAL INFORMATION:
; APPLICANT: Wang, Liangsu
; APPLICANT: Zamudio, Carlos
; APPLICANT: Malone, Cheryl
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari
; APPLICANT: Zyskind, Judith
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John
; APPLICANT: Carr, Grant
; APPLICANT: Yamamoto, Robert
; APPLICANT: Forsyth, R.
; APPLICANT: Xu, H.
; TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
; FILE REFERENCE: ELITRA.034A
; CURRENT APPLICATION NUMBER: US/10/282,122A
; CURRENT FILING DATE: 2003-02-20
```

```
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/230,335
; PRIOR FILING DATE: 2000-09-06
; PRIOR APPLICATION NUMBER: 60/230,347
; PRIOR FILING DATE: 2000-09-09
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/267,636
; PRIOR FILING DATE: 2001-02-09
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 78614
; SOFTWARE: Patentin version 3.1
; SEQ ID NO 48048
; LENGTH: 1129
; TYPE: PRT
; ORGANISM: Burkholderia cepacia
US-10-282-122A-48048

Query Match 10.9%; Score 84; DB 12; Length 1129;
Best Local Similarity 27.1%; Pred. No. 16;
Matches 46; Conservative 20; Mismatches 72; Indels 32; Gaps 7;

QY 8 AFAAIVVSGSALAGVWP-----QWGGGNHN--GGGNSGSDPTSILSIYQYG 51
Db 26 AQAGAAVPPGAHAGLSPPRIWTLAAAYLGLFQAGAAQYAGGGSATGGASSISV---81
QY 52 SANAALALQSDARKYDQLVTRVVTHEMAHAGQADNSTIELTQNGFRNNAT-IDQWNAKN 110
Db 82 -GNGSVATQVNSTAFGNLSSTAAGTSATA-LGPGAFAMGDGSTAVGNAQAATGVDSDASLGV 139
QY 111 SDITVQYGGNNAALVNQTASDSSVM-----VRQVGFNNATAN-QY 151
Db 140 QAIGSGAYSVAIGNLSSATQSGAVAMSGSAAATGVAAILGLNNAFASQCY 189

RESULT 12
US-09-810-264-28
; Sequence 28, Application US/09810264
; Patent No. US20020076775A1
; GENERAL INFORMATION:
; APPLICANT: Crane, Virginia C.
; APPLICANT: Famodu, Omolayo O.
; APPLICANT: Hu, Xu
; APPLICANT: Lu, Guihua
; APPLICANT: Zhang, Lingyu
; TITLE OF INVENTION: WRKY Transcription Factors and Methods
; TITLE OF INVENTION: of Use
; FILE REFERENCE: 1183
; CURRENT APPLICATION NUMBER: US/09/810,264
; CURRENT FILING DATE: 2001-03-16
; PRIOR APPLICATION NUMBER: US 60/190,467
; PRIOR FILING DATE: 2000-03-17
; NUMBER OF SEQ ID NOS: 43
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 28
; LENGTH: 278
; TYPE: PRT
; ORGANISM: Triticum aestivum
US-09-810-264-28

Query Match 10.8%; Score 83.5; DB 9; Length 278;
Best Local Similarity 20.8%; Pred. No. 3;
```



```
; PRIOR FILING DATE: 2001-04-23
; PRIOR APPLICATION NUMBER: US 60/287,066
; PRIOR FILING DATE: 2001-04-27
; PRIOR APPLICATION NUMBER: US 60/295,890
; PRIOR FILING DATE: 2001-06-05
; PRIOR APPLICATION NUMBER: US 60/303,899
; PRIOR FILING DATE: 2001-07-09
; PRIOR APPLICATION NUMBER: US 60/316,362
; PRIOR FILING DATE: 2001-08-31
; NUMBER OF SEQ ID NOS: 8603
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 3213
; LENGTH: 435
; TYPE: PRT
; ORGANISM: Aspergillus fumigatus
US-10-128-714-3213

Query Match      10.7%; Score 82.5; DB 14; Length 435;
Best Local Similarity 19.6%; Pred.No. 6.6;
Matches 31; Conservative 30; Mismatches 66; Indels 31; Gaps 6;

QY 13 VVSGSALAGVFPQWGGGNGHNGSSGPDSTLSIYQGSANAALALOSDARKYDQL--- 69
Db 274 VITGHKGSVTCVRWGGTGKI---YTSSHDRIKIWNAQGSLLQTLNAAHVRVNLALS 329

QY 70 ---VTRVVTHEMAHAGQ--GADNSTIELTQNGFR-----NNATIDQ-----WN 107
Db 330 TDFALRTAYHD--HTGKVPGSDEKVAVAKRFEQAAMVNNKIVEKLYSASDDFTMYLWD 387

QY 108 AKNSDITVQYGGNNAALVNQATSDSSVMVRQVPGNN 145
Db 388 PENSTKPIARLLGHQKEVNHVTFSPDMAYIASAGFDNH 425
```

Search completed: August 2, 2004, 15:36:11  
Job time : 36.8 secs

This Page Blank (uspto)

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: August 2, 2004, 14:48:33 ; Search time 167.9 Seconds

(without alignments)

877,809 Million cell updates/sec

Title: US-09-543-407-20

Perfect score: 774

Sequence: 1 MKLLKVAAPAAIWGSGSALA.....DSSVMVRQVGFNNATANYQ 151

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 6019581 seqs, 976053577 residues

Total number of hits satisfying chosen parameters: 6019581

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Pending Patents AA.Main:\*  
1: /cgn2\_6/prodata/2/paa/pctus\_COMB.pep.\*  
2: /cgn2\_6/prodata/2/paa/US06\_COMB.pep.\*  
3: /cgn2\_6/prodata/2/paa/US07\_COMB.pep.\*  
4: /cgn2\_6/prodata/2/paa/US08\_COMB.pep.\*  
5: /cgn2\_6/prodata/2/paa/US081\_COMB.pep.\*  
6: /cgn2\_6/prodata/2/paa/US082\_COMB.pep.\*  
7: /cgn2\_6/prodata/2/paa/US083\_COMB.pep.\*  
8: /cgn2\_6/prodata/2/paa/US084\_COMB.pep.\*  
9: /cgn2\_6/prodata/2/paa/US085\_COMB.pep.\*  
10: /cgn2\_6/prodata/2/paa/US086\_COMB.pep.\*  
11: /cgn2\_6/prodata/2/paa/US087\_COMB.pep.\*  
12: /cgn2\_6/prodata/2/paa/US088\_COMB.pep.\*  
13: /cgn2\_6/prodata/2/paa/US089\_COMB.pep.\*  
14: /cgn2\_6/prodata/2/paa/US090\_COMB.pep.\*  
15: /cgn2\_6/prodata/2/paa/US091\_COMB.pep.\*  
16: /cgn2\_6/prodata/2/paa/US092\_COMB.pep.\*  
17: /cgn2\_6/prodata/2/paa/US093\_COMB.pep.\*  
18: /cgn2\_6/prodata/2/paa/US094\_COMB.pep.\*  
19: /cgn2\_6/prodata/2/paa/US095\_COMB.pep.\*  
20: /cgn2\_6/prodata/2/paa/US096\_COMB.pep.\*  
21: /cgn2\_6/prodata/2/paa/US097A\_COMB.pep.\*  
22: /cgn2\_6/prodata/2/paa/US097B\_COMB.pep.\*  
23: /cgn2\_6/prodata/2/paa/US098\_COMB.pep.\*  
24: /cgn2\_6/prodata/2/paa/US099A\_COMB.pep.\*  
25: /cgn2\_6/prodata/2/paa/US099B\_COMB.pep.\*  
26: /cgn2\_6/prodata/2/paa/US100\_COMB.pep.\*  
27: /cgn2\_6/prodata/2/paa/US101\_COMB.pep.\*  
28: /cgn2\_6/prodata/2/paa/US102\_COMB.pep.\*  
29: /cgn2\_6/prodata/2/paa/US103\_COMB.pep.\*  
30: /cgn2\_6/prodata/2/paa/US104\_COMB.pep.\*  
31: /cgn2\_6/prodata/2/paa/US106\_COMB.pep.\*  
32: /cgn2\_6/prodata/2/paa/US107\_COMB.pep.\*  
33: /cgn2\_6/prodata/2/paa/US60\_COMB.pep.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No. Score Match Length DB ID Description

1	774	100.0	151	19	US-09-543-407-20	Sequence 20, Appl
2	709	91.6	151	19	US-09-543-407-26	Sequence 26, Appl
3	689	89.0	151	19	US-09-543-407-5	Sequence 5, Appl
4	684	88.4	151	6	US-08-233-642A-57	Sequence 57, Appl
5	662	85.5	151	19	US-09-543-407-28	Sequence 28, Appl
6	622	80.4	151	19	US-09-543-407-18	Sequence 18, Appl
7	619	80.0	151	19	US-09-543-407-22	Sequence 22, Appl
8	611	78.9	151	19	US-09-543-407-12	Sequence 12, Appl
9	609	78.7	151	19	US-09-543-407-14	Sequence 14, Appl
10	606	78.3	151	19	US-09-543-407-24	Sequence 24, Appl
11	602	77.8	151	19	US-09-543-407-31	Sequence 31, Appl
12	597	77.1	151	19	US-09-543-407-30	Sequence 30, Appl
13	574	74.2	151	19	US-09-543-407-16	Sequence 16, Appl
14	528	68.2	151	19	US-09-543-407-7	Sequence 7, Appl
15	525	67.8	151	13	US-08-978-878-4	Sequence 4, Appl
16	525	67.8	151	21	US-09-741-873B-4	Sequence 4, Appl
17	523	67.6	151	33	US-60-352-946-2	Sequence 2, Appl
18	523	67.6	151	33	US-60-444-371-2	Sequence 2, Appl
19	504	65.1	120	6	US-08-233-642A-55	Sequence 55, Appl
20	470	60.7	109	19	US-09-543-407-34	Sequence 34, Appl
21	466	60.2	158	16	US-09-252-691C-5834	Sequence 5834, Ap
22	466	60.2	158	16	US-09-252-691C-5834	Sequence 5834, Ap
23	466	60.2	158	30	US-10-417-886-5834	Sequence 5834, Ap
24	447	57.8	131	13	US-08-978-878-2	Sequence 2, Appl
25	447	57.8	131	21	US-09-741-873B-2	Sequence 2, Appl
26	343	47.3	109	19	US-09-543-407-35	Sequence 35, Appl
27	263	34.0	68	19	US-09-543-407-37	Sequence 37, Appl
28	166.5	21.5	70	19	US-09-543-407-32	Sequence 32, Appl
29	153	19.8	48	19	US-09-543-407-39	Sequence 39, Appl
30	105.5	13.6	145	21	US-09-739-449-8854	Sequence 8854, Ap
31	105.5	13.6	145	23	US-09-803-110-8854	Sequence 20638, A
32	101.5	13.1	445	23	US-10-369-493-20638	Sequence 20638, A
33	101.5	13.1	445	33	US-60-360-039-20638	Sequence 20638, A
34	99	12.8	492	21	US-09-708-427-7657	Sequence 7657, Ap
35	99	12.8	544	21	US-09-708-427-7656	Sequence 7656, Ap
36	99	12.8	573	21	US-09-708-427-7655	Sequence 7655, Ap
37	99	12.8	955	24	US-09-935-625-7946	Sequence 7946, Ap
38	99	12.8	955	24	US-09-935-625-25158	Sequence 25158, A
39	99	12.8	1036	24	US-09-935-625-7945	Sequence 7945, Ap
40	99	12.8	1036	24	US-09-935-625-25157	Sequence 25157, A
41	99	12.8	1055	19	US-09-570-581A-1851	Sequence 1851, Ap
42	99	12.8	1055	19	US-09-573-655A-349	Sequence 349, App
43	99	12.8	1055	19	US-09-573-655A-667	Sequence 667, App
44	99	12.8	1055	19	US-09-573-655B-349	Sequence 349, App
45	99	12.8	1055	19	US-09-573-655B-667	Sequence 667, App

ALIGNMENTS

RESULT 1  
US-09-543-407-20  
; Sequence 20, Application US/09543407  
; GENERAL INFORMATION:  
; APPLICANT: White, Aaron P.  
; APPLICANT: Doran, James L.  
; APPLICANT: Collinson, S. Karen  
; APPLICANT: Kay, William W.  
; TITLE OF INVENTION: BACTERIAL FIMBRIAL SYSTEM FOR PRESENTATION OF HETEROLOGOUS PEPTIDE SEQUENCES  
; FILE OF INVENTION: 920043-406  
; FILE REFERENCE: 920043-406  
; CURRENT APPLICATION NUMBER: US/09/543,407  
; CURRENT FILING DATE: 2000-04-05  
; NUMBER OF SEQ ID NOS: 59  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 20  
; TYPE: PRT  
; LENGTH: 151  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: Recombinant Salmonella enteritidis 3b afgA  
; OTHER INFORMATION: sequence containing the replacement fragment  
; OTHER INFORMATION: encoding PT3 from GP63 of Leishmania major.

## US-09-543-407-20

Query Match 100.0%; Score 774; DB 19; Length 151;  
 Best Local Similarity 100.0%; Pred. No. 3.3e-75;  
 Matches 151; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MLLKVAFAAIVVSGSALAGVVPQWGGGNGHNGGNSGPDSTLSIYQYGSANAALALQ 60

DB 1 MLLKVAFAAIVVSGSALAGVVPQWGGGNGHNGGNSGPDSTLSIYQYGSANAALALQ 60

QY 61 SDARKYDQVTRVVTTHMAHAGGADNSTIELTQNGFRNNATIDOWNAKNSDITVGOYGG 120

DB 61 SDARKYDQVTRVVTTHMAHAGGADNSTIELTQNGFRNNATIDOWNAKNSDITVGOYGG 120

QY 121 NNAALVNQTSASSVMVRQVGFNNATANQY 151

DB 121 NNAALVNQTSASSVMVRQVGFNNATANQY 151

## RESULT 2

US-09-543-407-26

; Sequence 26, Application US/09543407

; GENERAL INFORMATION:

; APPLICANT: White, Aaron P.

; APPLICANT: Doran, James L.

; APPLICANT: Collinson, S. Karen

; APPLICANT: Kay, William W.

; TITLE OF INVENTION: BACTERIAL FIMBRIAL SYSTEM FOR

; TITLE OF INVENTION: PRESENTATION OF HETEROLOGOUS PEPTIDE SEQUENCES

; FILE REFERENCE: 920043.406

; CURRENT APPLICATION NUMBER: US/09/543,407

; CURRENT FILING DATE: 2000-04-05

; NUMBER OF SEQ ID NOS: 59

; SOFTWARE: FastSeq for Windows Version 4.0

; SEQ ID NO 26

; LENGTH: 151

; TYPE: PRT

; ORGANISM: Artificial Sequence

; FEATURE:

; OTHER INFORMATION: Recombinant Salmonella enteritidis 3b afgA

; OTHER INFORMATION: sequence containing the replacement fragment

; OTHER INFORMATION: encoding P73 from Gp63 of Leishmania major.

US-09-543-407-26

Query Match 91.6%; Score 709; DB 19; Length 151;  
 Best Local Similarity 89.9%; Pred. No. 3.8e-68;  
 Matches 143; Conservative 0; Mismatches 0; Indels 16; Gaps 2;

QY 1 MLLKVAFAAIVVSGSALAGVVPQWGGGNGHNGGNSGPDSTLSIYQYGSANAALALQ 60

DB 1 MLLKVAFAAIVVSGSALAGVVPQWGGGNGHNGGNSGPDSTLSIYQYGSANAALALQ 60

QY 61 SDARKYDQVTRVVTTHMAHAGGADNSTIELTQNGFRNNATIDOWNAKNSD 112

DB 58 -----YDQLVTRVVTTHMAHAGGADNSTIELTQNGFRNNATIDOWNAKNSD 112

QY 113 ITVGOYGENNAALVNQTSASSVMVRQVGFNNATANQY 151

DB 113 ITVGOYGENNAALVNQTSASSVMVRQVGFNNATANQY 151

## RESULT 3

US-09-543-407-5

; Sequence 5, Application US/09543407

; GENERAL INFORMATION:

; APPLICANT: White, Aaron P.

; APPLICANT: Doran, James L.

; APPLICANT: Collinson, S. Karen

; APPLICANT: Kay, William W.

; TITLE OF INVENTION: BACTERIAL FIMBRIAL SYSTEM FOR

; TITLE OF INVENTION: PRESENTATION OF HETEROLOGOUS PEPTIDE SEQUENCES

; FILE REFERENCE: 920043.406

; CURRENT APPLICATION NUMBER: US/09/543,407

; CURRENT FILING DATE: 2000-04-05

; NUMBER OF SEQ ID NOS: 59

; SOFTWARE: FastSeq for Windows Version 4.0

; SEQ ID NO 5

; LENGTH: 151

; TYPE: PRT

; ORGANISM: Salmonella enteritidis

US-09-543-407-5

Query Match 89.0%; Score 689; DB 19; Length 151;  
 Best Local Similarity 90.7%; Pred. No. 5.6e-66;  
 Matches 137; Conservative 3; Mismatches 11; Indels 0; Gaps 0;

QY 1 MLLKVAFAAIVVSGSALAGVVPQWGGGNGHNGGNSGPDSTLSIYQYGSANAALALQ 60

DB 1 MLLKVAFAAIVVSGSALAGVVPQWGGGNGHNGGNSGPDSTLSIYQYGSANAALALQ 60

QY 61 SDARKYDQVTRVVTTHMAHAGGADNSTIELTQNGFRNNATIDOWNAKNSDITVGOYGG 120

DB 61 SDARKSETTITQSGYNGADVGGADNSTIELTQNGFRNNATIDOWNAKNSDITVGOYGG 120

QY 121 NNAALVNQTSASSVMVRQVGFNNATANQY 151

DB 121 NNAALVNQTSASSVMVRQVGFNNATANQY 151

## RESULT 4

US-08-233-642A-57

; Sequence 57, Application US/08233642A

; GENERAL INFORMATION:

; APPLICANT: Kay, William W.

; APPLICANT: Collinson, S. Karen

; APPLICANT: Clouthier, Sharon C.

; APPLICANT: Doran, James L.

; TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR SALMONELLA-

; TITLE OF INVENTION: BASED VACCINES

; NUMBER OF SEQUENCES: 58

; NUMBER OF SEQUENCES: -

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Seed and Berry

; STREET: 6300 Columbia Center, 701 Fifth Avenue

; CITY: Seattle

; STATE: Washington

; COUNTRY: U.S.A.

; ZIP: 98104-7092

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: Patent In Release #1.0, Version #1.25

; CURRENT APPLICATION NUMBER: US/08/233,642A

; FILING DATE: 26-APR-1994

; CLASSIFICATION: 424

; ATTORNEY/AGENT INFORMATION:

; NAME: King, Joshua

; REGISTRATION NUMBER: 35,570

; REFERENCE/DOCKET NUMBER: 920043.403C3

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: (206) 622-4900

; TELEFAX: (206) 682-6031

; TELETYPE: 3723836 SEEDANBERRY

; INFORMATION FOR SEQ ID NO: 57:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 151 amino acids

; TYPE: amino acid

; TOPOLOGY: linear

; MOLECULE TYPE: protein

US-08-233-642A-57

Query Match 88.4%; Score 684; DB 6; Length 151;  
 Best Local Similarity 90.1%; Pred. No. 1.9e-65;  
 Matches 136; Conservative 3; Mismatches 12; Indels 0; Gaps 0;

```

QY 1 MLLKVAFAAIVVSGSALAGVVPQWGGGNGHNGGSSGPDSTLSIYQGSANAALALQ 60
Db 1 MLLKVAFAAIVVSGSALAGVVPQWGGGNGHNGGSSGPDSTLSIYQGSANAALALQ 60
QY 61 SDARKYDQLVTRVVTHEMAHAGGADNSTIETQNGFRNNATIDQWNAKNSDITVGYGG 120
Db 61 SDARKSETTITQSGYNGADVGQGDNSTIETQNGFRNNATIDQWNAKNSDITVGYGG 120
QY 121 NNAALVNQASDSSVMVRQVGFNNATANQY 151
Db 121 NNPAALVNQASDSSVMVRQVGFNNATANQY 151

```

## RESULT 5

```

US-09-543-407-28
; Sequence 28, Application US/09543407
; GENERAL INFORMATION:
; APPLICANT: White, Aaron P.
; APPLICANT: Doran, James L.
; APPLICANT: Collinson, S. Karen
; APPLICANT: Kay, William W.
; TITLE OF INVENTION: BACTERIAL FIMBRIAL SYSTEM FOR
; PRESENTATION OF HETEROLOGOUS PEPTIDE SEQUENCES
; FILE REFERENCE: 920043.406
; CURRENT FILING DATE: 2000-04-05
; NUMBER OF SEQ ID NOS: 59
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 28
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Recombinant Salmonella enteritidis 3b afga
; OTHER INFORMATION: sequence containing the replacement fragment
; OTHER INFORMATION: encoding PT3 from GP63 of Leishmania major.
US-09-543-407-28

```

```

Query Match 85.5%; Score 662; DB 19; Length 151;
Best Local Similarity 81.9%; Pred. No. 4.7e-63;
Matches 136; Conservative 0; Mismatches 0; Indels 30; Gaps 2;

QY 1 MLLKVAFAAIVVSGSALAGVVPQWGGGNGHNGGSSGPDSTLSIYQGSANAALALQ 60
Db 1 MLLKVAFAAIVVSGSALAGVVPQWGGGNGHNGGSSGPDSTLSIYQGSANAALALQ 60
QY 61 SDARK-----YDQLVTRVVTHEMAHAGGADNSTIETQNGFRNNATIDQ 105
Db 61 SDARKSETTITQSGYNGADYDQLVTRVVTHEMAHA-----FNNATIDQ 105
QY 106 WNAKNSDITVGYGGNNAALVNQASDSSVMVRQVGFNNATANQY 151
Db 106 WNAKNSDITVGYGGNNAALVNQASDSSVMVRQVGFNNATANQY 151

```

## RESULT 6

```

US-09-543-407-18
; Sequence 18, Application US/09543407
; GENERAL INFORMATION:
; APPLICANT: White, Aaron P.
; APPLICANT: Doran, James L.
; APPLICANT: Collinson, S. Karen
; APPLICANT: Kay, William W.
; TITLE OF INVENTION: BACTERIAL FIMBRIAL SYSTEM FOR
; PRESENTATION OF HETEROLOGOUS PEPTIDE SEQUENCES
; FILE REFERENCE: 920043.406
; CURRENT FILING DATE: 2000-04-05
; NUMBER OF SEQ ID NOS: 59
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 18
; LENGTH: 151

```

```

; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Recombinant Salmonella enteritidis 3b afga
; OTHER INFORMATION: sequence containing the replacement fragment
; OTHER INFORMATION: encoding PT3 from GP63 of Leishmania major.
US-09-543-407-18

```

```

Query Match 80.4%; Score 622; DB 19; Length 151;
Best Local Similarity 73.6%; Pred. No. 1e-58;
Matches 128; Conservative 0; Mismatches 0; Indels 46; Gaps 2;

QY 1 MLLKVAFAAIVVSGSALAGVVPQWGGGNGHNGGSSGPDSTLSIYQGSANAALALQ 60
Db 1 MLLKVAFAAIVVSGSALAGVVPQWGGGNGHNGGSSGPD----- 42
QY 61 SDARKYDQLVTRVVTHEMAHA-----GQADNSTIETQNGF 97
Db 43 -----YDQLVTRVVTHEMAHALQSDARKSETTITQSGYNGADVGQGDNSTIETQNGF 97
QY 98 RNNATIDQWNAKNSDITVGYGGNNAALVNQASDSSVMVRQVGFNNATANQY 151
Db 98 RNNATIDQWNAKNSDITVGYGGNNAALVNQASDSSVMVRQVGFNNATANQY 151

```

## RESULT 7

```

US-09-543-407-22
; Sequence 22, Application US/09543407
; GENERAL INFORMATION:
; APPLICANT: White, Aaron P.
; APPLICANT: Doran, James L.
; APPLICANT: Collinson, S. Karen
; APPLICANT: Kay, William W.
; TITLE OF INVENTION: BACTERIAL FIMBRIAL SYSTEM FOR
; PRESENTATION OF HETEROLOGOUS PEPTIDE SEQUENCES
; FILE REFERENCE: 920043.406
; CURRENT FILING DATE: 2000-04-05
; NUMBER OF SEQ ID NOS: 59
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 22
; LENGTH: 151
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Recombinant Salmonella enteritidis 3b afga
; OTHER INFORMATION: sequence containing the replacement fragment
; OTHER INFORMATION: encoding PT3 from GP63 of Leishmania major.
US-09-543-407-22

```

```

Query Match 80.0%; Score 619; DB 19; Length 151;
Best Local Similarity 74.6%; Pred. No. 2.2e-58;
Matches 129; Conservative 0; Mismatches 0; Indels 44; Gaps 2;

QY 1 MLLKVAFAAIVVSGSALAGVVPQWGGGNGHNGGSSGPDSTLSIYQGSANAALALQ 60
Db 1 MLLKVAFAAIVVSGSALAGVVPQWGGGNGHNGGSSGPDSTLSIYQGSANAALALQ 60
QY 61 SDARK-----YDQLVTRVVTHEMAHAGGADNSTIETQNGFR 98
Db 61 SDARKSETTITQSGYNGADVGQGDNDYDQLVTRVVTHEMAHA----- 103
QY 99 NNATIDQWNAKNSDITVGYGGNNAALVNQASDSSVMVRQVGFNNATANQY 151
Db 104 -----DQWNAKNSDITVGYGGNNAALVNQASDSSVMVRQVGFNNATANQY 151

```

## RESULT 8

```

US-09-543-407-12
; Sequence 12, Application US/09543407
; GENERAL INFORMATION:
; APPLICANT: White, Aaron P.
; APPLICANT: Doran, James L.

```

APPLICANT: Collinson, S. Karen  
APPLICANT: Kay, William W.  
TITLE OF INVENTION: BACTERIAL FIMBRIAL SYSTEM FOR  
FILE REFERENCE: 920043.406  
CURRENT APPLICATION NUMBER: US/09/543,407  
CURRENT FILING DATE: 2000-04-05  
NUMBER OF SEQ ID NOS: 59  
SOFTWARE: FastSeq for Windows Version 4.0  
SEQ ID NO 12  
LENGTH: 151  
TYPE: PRT  
ORGANISM: Artificial Sequence  
FEATURE:  
OTHER INFORMATION: Recombinant Salmonella enteritidis 3b afgA  
OTHER INFORMATION: sequence containing the replacement fragment  
OTHER INFORMATION: encoding PT3 from GP63 of Leishmania major.  
US-09-543-407-12

Query Match 78.9%; Score 611; DB 19; Length 151;  
Best Local Similarity 80.8%; Pred. No. 1.6e-57;  
Matches 122; Conservative 7; Mismatches 22; Indels 0; Gaps 0;

QY 1 MKLLKVAFAAIVVSGSALAGVVPQWGGGNGHNGGNSGPDSTLSIYQYGSANAALALQ 60  
DB 1 MKLLKVAFAAIVVSGSALAGVVPQWGGGNGHNGGNSGPDSTLSIYQYGSANAALALQ 60  
QY 61 SDARKYDQLVTRVVTHEMAHAGGADNSTIELTQNGFRNNATIDQWNAKNSDITVGYGG 120  
DB 61 SDARKSETTITQSGYNGADVGCGADNSTIELTQNGFRNNATIDQWNAKNSDITVGYGG 120  
QY 121 NNAALVNOTASDSSVMVRQVGFNNATANQY 151  
DB 121 NNAALVNYDQLVTRVVTHEMAHANNATANQY 151

RESULT 9  
US-09-543-407-14  
Sequence 14, Application US/09543407  
GENERAL INFORMATION:  
APPLICANT: White, Aaron P.  
APPLICANT: Doran, James L.  
APPLICANT: Collinson, S. Karen  
APPLICANT: Kay, William W.  
TITLE OF INVENTION: BACTERIAL FIMBRIAL SYSTEM FOR  
FILE REFERENCE: 920043.406  
CURRENT APPLICATION NUMBER: US/09/543,407  
CURRENT FILING DATE: 2000-04-05  
NUMBER OF SEQ ID NOS: 59  
SOFTWARE: FastSeq for Windows Version 4.0  
SEQ ID NO 14  
LENGTH: 151  
TYPE: PRT  
ORGANISM: Artificial Sequence  
FEATURE:  
OTHER INFORMATION: Recombinant Salmonella enteritidis 3b afgA  
OTHER INFORMATION: sequence containing the replacement fragment  
OTHER INFORMATION: encoding PT3 from GP63 of Leishmania major.  
US-09-543-407-14

Query Match 78.7%; Score 609; DB 19; Length 151;  
Best Local Similarity 81.5%; Pred. No. 2.7e-57;  
Matches 123; Conservative 6; Mismatches 22; Indels 0; Gaps 0;

QY 1 MKLLKVAFAAIVVSGSALAGVVPQWGGGNGHNGGNSGPDSTLSIYQYGSANAALALQ 60  
DB 1 MKLLKVAFAAIVVSGSALAGVVPQWGGGNGHNGGNSGPDSTLSIYQYGSANAALALQ 60  
QY 61 SDARKYDQLVTRVVTHEMAHAGGADNSTIELTQNGFRNNATIDQWNAKNSDITVGYGG 120  
DB 61 SDARKSETTITQSGYNGADVGCGADNSTIELTQNGFRNNATIDQWNAKNSDITVGYDQ 120

QY 121 NNAALVNOTASDSSVMVRQVGFNNATANQY 151  
DB 121 LVTRVVTHEMAHASVMVRQVGFNNATANQY 151

RESULT 10  
US-09-543-407-24  
Sequence 24, Application US/09543407  
GENERAL INFORMATION:  
APPLICANT: White, Aaron P.  
APPLICANT: Doran, James L.  
APPLICANT: Collinson, S. Karen  
APPLICANT: Kay, William W.  
TITLE OF INVENTION: BACTERIAL FIMBRIAL SYSTEM FOR  
FILE REFERENCE: 920043.406  
CURRENT APPLICATION NUMBER: US/09/543,407  
CURRENT FILING DATE: 2000-04-05  
NUMBER OF SEQ ID NOS: 59  
SOFTWARE: FastSeq for Windows Version 4.0  
SEQ ID NO 24  
LENGTH: 151  
TYPE: PRT  
ORGANISM: Artificial Sequence  
FEATURE:  
OTHER INFORMATION: Recombinant Salmonella enteritidis 3b afgA  
OTHER INFORMATION: sequence containing the replacement fragment  
OTHER INFORMATION: encoding PT3 from GP63 of Leishmania major.  
US-09-543-407-24

Query Match 78.3%; Score 606; DB 19; Length 151;  
Best Local Similarity 82.1%; Pred. No. 5.7e-57;  
Matches 124; Conservative 5; Mismatches 22; Indels 0; Gaps 0;

QY 1 MKLLKVAFAAIVVSGSALAGVVPQWGGGNGHNGGNSGPDSTLSIYQYGSANAALALQ 60  
DB 1 MKLLKVAFAAIVVSGSALAGVVPQWGGGNGHNGGNSGPDSTLSIYQYGSANAALALQ 60  
QY 61 SDARKYDQLVTRVVTHEMAHAGGADNSTIELTQNGFRNNATIDQWNAKNSDITVGYGG 120  
DB 61 SDARKSETTITQSGYNGADVGCGADNSTIELTQNGFRNNATIDQWNAKNSDITVTRVT 120  
QY 121 NNAALVNOTASDSSVMVRQVGFNNATANQY 151  
DB 121 HEMAHANQTASDSSVMVRQVGFNNATANQY 151

RESULT 11  
US-09-543-407-31  
Sequence 31, Application US/09543407  
GENERAL INFORMATION:  
APPLICANT: White, Aaron P.  
APPLICANT: Doran, James L.  
APPLICANT: Collinson, S. Karen  
APPLICANT: Kay, William W.  
TITLE OF INVENTION: BACTERIAL FIMBRIAL SYSTEM FOR  
FILE REFERENCE: 920043.406  
CURRENT APPLICATION NUMBER: US/09/543,407  
CURRENT FILING DATE: 2000-04-05  
NUMBER OF SEQ ID NOS: 59  
SOFTWARE: FastSeq for Windows Version 4.0  
SEQ ID NO 31  
LENGTH: 131  
TYPE: PRT  
ORGANISM: Salmonella enteritidis  
US-09-543-407-31

Query Match 77.8%; Score 602; DB 19; Length 131;  
Best Local Similarity 89.3%; Pred. No. 1.3e-56;  
Matches 117; Conservative 3; Mismatches 11; Indels 0; Gaps 0;

QY 21 GWVPQWGGGNGHNGGNSGPDSTLSIYQYGSANAALALQSDARKYDQLVTRVVTHEMAH 80



```
Db 1 GVVPQWGGGNGGNGSSGPDSTLSIYQGSANAALALQSDARKSETTITQSGYNGAD 60
QY 81 ACGADNSTIETQNGFRNATIDOWNAKNSDITVGYGGNNAALVNQTSDDSSVMVROV 140
Db 61 VCGADNSTIETQNGFRNATIDOWNAKNSDITVGYGGNNAALVNQTSDDSSVMVROV 120
QY 141 GFGNNATANQY 151
Db 121 GFGNNATANQY 131

RESULT 12
US-09-543-407-30
; Sequence 30, Application US/09543407
; GENERAL INFORMATION:
; APPLICANT: White, Aaron P.
; APPLICANT: Doran, James L.
; APPLICANT: Collinson, S. Karen
; TITLE OF INVENTION: BACTERIAL FIMBRIAL SYSTEM FOR
; FILE OF INVENTION: PRESENTATION OF HETEROLOGOUS PEPTIDE SEQUENCES
; FILE REFERENCE: 920043.406
; CURRENT APPLICATION NUMBER: US/09/543,407
; CURRENT FILING DATE: 2000-04-05
; NUMBER OF SEQ ID NOS: 59
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 30
; LENGTH: 151
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Recombinant Salmonella enteritidis 3b afgA
; OTHER INFORMATION: sequence containing the replacement fragment
; OTHER INFORMATION: encoding PT3 from GP63 of Leishmania major.
US-09-543-407-30

Query Match 77.1%; Score 597; DB 19; Length 151;
Best Local Similarity 81.5%; Pred. No. 5.4e-56;
Matches 123; Conservative 5; Mismatches 23; Indels 0; Gaps 0;

QY 1 MKLLKVAFAAIVVSGSALAGVVPQWGGGNGGNGSSGPDSTLSIYQGSANAALALQ 60
Db 1 MKLLKVAFAAIVVSGSALAGVVPQWGGGNGGNGSSGPDSTLSIYQGSANAALALQ 60
QY 61 SDARKYDQLVTRVVTHEMAHAGGADNSTIETQNGFRNATIDOWNAKNSDITVGYGG 120
Db 61 SDARKSETTITQSGYNGADVGQGDNSTIETQNGFRNATIDQVLTVRVTHEMAHAGG 120
QY 121 NNAALVNQTSDDSSVMVROVQFGNNATANQY 151
Db 121 NNAALVNQTSDDSSVMVROVQFGNNATANQY 151

RESULT 13
US-09-543-407-16
; Sequence 16, Application US/09543407
; GENERAL INFORMATION:
; APPLICANT: White, Aaron P.
; APPLICANT: Doran, James L.
; APPLICANT: Collinson, S. Karen
; TITLE OF INVENTION: BACTERIAL FIMBRIAL SYSTEM FOR
; FILE OF INVENTION: PRESENTATION OF HETEROLOGOUS PEPTIDE SEQUENCES
; FILE REFERENCE: 920043.406
; CURRENT APPLICATION NUMBER: US/09/543,407
; CURRENT FILING DATE: 2000-04-05
; NUMBER OF SEQ ID NOS: 59
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 16
; LENGTH: 151
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Recombinant Salmonella enteritidis 3b afgA
; OTHER INFORMATION: sequence containing the replacement fragment
; OTHER INFORMATION: encoding PT3 from GP63 of Leishmania major.
US-09-543-407-16

Query Match 77.1%; Score 597; DB 19; Length 151;
Best Local Similarity 81.5%; Pred. No. 5.4e-56;
Matches 123; Conservative 5; Mismatches 23; Indels 0; Gaps 0;

QY 1 MKLLKVAFAAIVVSGSALAGVVPQWGGGNGGNGSSGPDSTLSIYQGSANAALALQ 60
Db 1 MKLLKVAFAAIVVSGSALAGVVPQWGGGNGGNGSSGPDSTLSIYQGSANAALALQ 60
QY 61 SDARKYDQLVTRVVTHEMAHAGGADNSTIETQNGFRNATIDOWNAKNSDITVGYGG 120
Db 61 SDARKSETTITQSGYNGADVGQGDNSTIETQNGFRNATIDQVLTVRVTHEMAHAGG 120
QY 121 NNAALVNQTSDDSSVMVROVQFGNNATANQY 151
Db 121 NNAALVNQTSDDSSVMVROVQFGNNATANQY 151

RESULT 14
US-09-543-407-7
; Sequence 7, Application US/09543407
; GENERAL INFORMATION:
; APPLICANT: White, Aaron P.
; APPLICANT: Doran, James L.
; APPLICANT: Collinson, S. Karen
; APPLICANT: Kay, William W.
; TITLE OF INVENTION: BACTERIAL FIMBRIAL SYSTEM FOR
; FILE OF INVENTION: PRESENTATION OF HETEROLOGOUS PEPTIDE SEQUENCES
; FILE REFERENCE: 920043.406
; CURRENT APPLICATION NUMBER: US/09/543,407
; CURRENT FILING DATE: 2000-04-05
; NUMBER OF SEQ ID NOS: 59
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 7
; LENGTH: 151
; TYPE: PRT
; ORGANISM: Escherichia coli
US-09-543-407-7

Query Match 68.2%; Score 528; DB 19; Length 151;
Best Local Similarity 68.9%; Pred. No. 1.6e-48;
Matches 104; Conservative 20; Mismatches 27; Indels 0; Gaps 0;

QY 1 MKLLKVAFAAIVVSGSALAGVVPQWGGGNGGNGSSGPDSTLSIYQGSANAALALQ 60
Db 1 MKLLKVAFAAIVVSGSALAGVVPQWGGGNGGNGSSGPDSTLSIYQGSANAALALQ 60
QY 61 SDARKYDQLVTRVVTHEMAHAGGADNSTIETQNGFRNATIDOWNAKNSDITVGYGG 120
Db 61 TDARNSDLTITQHGNGGADVGQGDSDSDITLQTFGNSATLQWNGKSEMTVKQFGG 120
QY 121 NNAALVNQTSDDSSVMVROVQFGNNATANQY 151
Db 121 GNGAAVDQTSNSSVNVTVQVFGNNATAHQY 151

RESULT 15
US-08-978-878-4
; Sequence 4, Application US/08978878
; GENERAL INFORMATION:
; APPLICANT: NORMARK, Staffan
; APPLICANT: OLSEN, Arne
; TITLE OF INVENTION: FIBRONECTIN BINDING PROTEIN AS WELL AS ITS PREPARATION
; FILE REFERENCE: 012889-081
; CURRENT APPLICATION NUMBER: US/08/978,878
; CURRENT FILING DATE: 1997-11-26
; EARLIER APPLICATION NUMBER: SE 8801723-1
; EARLIER FILING DATE: 1988-05-06
```

```
; FEATURE:
; OTHER INFORMATION: Recombinant Salmonella enteritidis 3b afgA
; OTHER INFORMATION: sequence containing the replacement fragment
; OTHER INFORMATION: encoding PT3 from GP63 of Leishmania major.
US-09-543-407-16

Query Match 74.2%; Score 574; DB 19; Length 151;
Best Local Similarity 80.8%; Pred. No. 1.7e-53;
Matches 122; Conservative 6; Mismatches 23; Indels 0; Gaps 0;

QY 1 MKLLKVAFAAIVVSGSALAGVVPQWGGGNGGNGSSGPDSTLSIYQGSANAALALQ 60
Db 1 MKLLKVAFAAIVVSGSALAGVVDQVLTVRVTHEMAHAGGPDSTLSIYQGSANAALALQ 60
QY 61 SDARKYDQLVTRVVTHEMAHAGGADNSTIETQNGFRNATIDOWNAKNSDITVGYGG 120
Db 61 SDARKSETTITQSGYNGADVGQGDNSTIETQNGFRNATIDQWNAKNSDITVGYGG 120
QY 121 NNAALVNQTSDDSSVMVROVQFGNNATANQY 151
Db 121 NNAALVNQTSDDSSVMVROVQFGNNATANQY 151

RESULT 14
US-09-543-407-7
; Sequence 7, Application US/09543407
; GENERAL INFORMATION:
; APPLICANT: White, Aaron P.
; APPLICANT: Doran, James L.
; APPLICANT: Collinson, S. Karen
; APPLICANT: Kay, William W.
; TITLE OF INVENTION: BACTERIAL FIMBRIAL SYSTEM FOR
; FILE OF INVENTION: PRESENTATION OF HETEROLOGOUS PEPTIDE SEQUENCES
; FILE REFERENCE: 920043.406
; CURRENT APPLICATION NUMBER: US/09/543,407
; CURRENT FILING DATE: 2000-04-05
; NUMBER OF SEQ ID NOS: 59
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 7
; LENGTH: 151
; TYPE: PRT
; ORGANISM: Escherichia coli
US-09-543-407-7

Query Match 68.2%; Score 528; DB 19; Length 151;
Best Local Similarity 68.9%; Pred. No. 1.6e-48;
Matches 104; Conservative 20; Mismatches 27; Indels 0; Gaps 0;

QY 1 MKLLKVAFAAIVVSGSALAGVVPQWGGGNGGNGSSGPDSTLSIYQGSANAALALQ 60
Db 1 MKLLKVAFAAIVVSGSALAGVVPQWGGGNGGNGSSGPDSTLSIYQGSANAALALQ 60
QY 61 SDARKYDQLVTRVVTHEMAHAGGADNSTIETQNGFRNATIDOWNAKNSDITVGYGG 120
Db 61 TDARNSDLTITQHGNGGADVGQGDSDSDITLQTFGNSATLQWNGKSEMTVKQFGG 120
QY 121 NNAALVNQTSDDSSVMVROVQFGNNATANQY 151
Db 121 GNGAAVDQTSNSSVNVTVQVFGNNATAHQY 151

RESULT 15
US-08-978-878-4
; Sequence 4, Application US/08978878
; GENERAL INFORMATION:
; APPLICANT: NORMARK, Staffan
; APPLICANT: OLSEN, Arne
; TITLE OF INVENTION: FIBRONECTIN BINDING PROTEIN AS WELL AS ITS PREPARATION
; FILE REFERENCE: 012889-081
; CURRENT APPLICATION NUMBER: US/08/978,878
; CURRENT FILING DATE: 1997-11-26
; EARLIER APPLICATION NUMBER: SE 8801723-1
; EARLIER FILING DATE: 1988-05-06
```

```
; EARLIER APPLICATION NUMBER: US 07/347,189
; EARLIER FILING DATE: 1989-05-04
; EARLIER APPLICATION NUMBER: US 07/789,437
; EARLIER FILING DATE: 1991-11-06
; EARLIER APPLICATION NUMBER: US 07/970,846
; EARLIER FILING DATE: 1992-11-03
; EARLIER APPLICATION NUMBER: US 08/187,865
; EARLIER FILING DATE: 1994-01-28
; EARLIER APPLICATION NUMBER: US 08/318,519
; EARLIER FILING DATE: 1994-10-05
; EARLIER APPLICATION NUMBER: US 08/495,959
; EARLIER FILING DATE: 1995-06-28
; NUMBER OF SEQ ID NOS: 10
; SOFTWARE: Patentin Ver. 2.0
; SEQ ID NO 4
; LENGTH: 151
; TYPE: PRT
; ORGANISM: Escherichia coli
US-08-978-878-4

Query Match      67.8%; Score 525; DB 13; Length 151;
Best Local Similarity 68.2%; Pred. No. 3.5e-48;
Matches 103; Conservative 21; Mismatches 27; Indels 0; Gaps 0;

QY      1 MKLLKVAAPAAIVVSGSALAGVVPQWGGGNGHGGNSSGPDSTLSIYQYGSANAALALQ 60
DB      1 MKLLKVAATAAIVFSGSAGVVPQYGGGNGHGGNNSGPNSELNIYQYGGNSALALQ 60

QY      61 SDARKYDQLVTRVVVTHEMAHAGGADNSTIELTQNGFRNNATIDOWNAKNSDITVQYGG 120
DB      61 TDARNSDLTITQHGCGGNGADVGGGSDSSIDLTLQRCFGNSATLDQWNGKNSMTYKQFGG 120

QY      121 NNAALVNQTASDSSVMVRQVFGNNATANQY 151
DB      121 GNGAAVDQTASNSNVNVTQVFGNNATAHQY 151
```

Search completed: August 2, 2004, 15:26:44  
Job time : 169.9 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: August 2, 2004, 14:49:38 ; Search time 17.8 Seconds

(without alignments)  
888.146 Million cell updates/sec

Title: US-09-543-407-20

Perfect score: 774

Sequence: 1 MKLLKVAAPAAIVVSGSALA.....DSVMVRQVGFNNATANYQ 151

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 601315 seqs, 104695340 residues

Total number of hits satisfying chosen parameters: 601315

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

Pending Patents AA New.\*  
1: /cgn2\_6/ptodata/2/paa/PCT\_NEW\_COMB.pep.\*  
2: /cgn2\_6/ptodata/2/paa/US06\_NEW\_COMB.pep.\*  
3: /cgn2\_6/ptodata/2/paa/US07\_NEW\_COMB.pep.\*  
4: /cgn2\_6/ptodata/2/paa/US08\_NEW\_COMB.pep.\*  
5: /cgn2\_6/ptodata/2/paa/US09\_NEW\_COMB.pep.\*  
6: /cgn2\_6/ptodata/2/paa/US10\_NEW\_COMB.pep.\*  
7: /cgn2\_6/ptodata/2/paa/US60\_NEW\_COMB.pep.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match %	Length	ID	Description
1	525	67.8	151	5	US-09-741-873C-4
2	447	57.8	131	5	US-09-741-873C-2
3	87	11.2	573	7	US-60-565-632-7907
4	87	11.2	573	7	US-60-579-062-7907
5	85	11.0	147	6	US-10-425-115-193207
6	85	11.0	412	7	US-60-565-632-7905
7	85	11.0	412	7	US-60-579-062-7905
8	85	11.0	841	7	US-60-565-632-7906
9	85	11.0	841	7	US-60-579-062-7906
10	85	11.0	2663	7	US-60-579-902-7377
11	81.5	10.5	436	7	US-60-581-351-7636
12	80.5	10.4	511	6	US-10-425-115-320950
13	80	10.3	581	1	PCT-US04-07412-1695
14	80	10.3	581	6	US-10-389-559-1695
15	79.5	10.3	321	7	US-60-566-841-11319
16	79	10.2	374	1	PCT-US04-11210-37
17	79	10.2	719	5	US-09-248-796A-17559
18	78.5	10.1	753	6	US-10-170-205E-35514
19	78.5	10.1	753	6	US-10-170-205E-35515
20	77	9.9	545	7	US-60-565-632-7973
21	77	9.9	545	7	US-60-579-062-7973
22	76.5	9.9	179	6	US-10-767-701-35342
23	76.5	9.9	234	1	PCT-US04-16242-10
24	76	9.8	556	6	US-10-425-115-337674
25	76	9.8	892	5	US-09-852-267B-5
26	76	9.8	892	6	US-10-872-768-5

27 76 9.8 892 6 US-10-872-769-5 Sequence 5, Appli  
28 76 9.8 1095 1 PCT-US03-24982A-273 Sequence 273, App  
29 76 9.8 1459 6 US-10-482-706-129 Sequence 129, App  
30 75.5 9.8 167 6 US-10-425-115-242035 Sequence 242035,  
31 75.5 9.8 276 6 US-10-425-115-339993 Sequence 339993,  
32 75.5 9.8 415 5 US-09-490-324-280 Sequence 280, App  
33 75.5 9.8 544 7 US-60-563-858-16 Sequence 16, Appl  
34 75.5 9.8 619 1 PCT-US04-10229-38 Sequence 38, Appl  
35 75 9.7 389 7 US-60-556-841-11867 Sequence 11867, A  
36 75 9.7 552 5 US-09-248-796A-25337 Sequence 25337, A  
37 74.5 9.6 179 6 US-10-425-115-346132 Sequence 346132,  
38 74.5 9.6 298 7 US-60-565-632-8923 Sequence 8923, Ap  
39 74.5 9.6 298 7 US-60-579-062-8923 Sequence 8923, Ap  
40 74.5 9.6 634 7 US-60-581-351-9813 Sequence 9813, Ap  
41 74.5 9.6 672 1 PCT-US04-07036-553 Sequence 553, App  
42 74 9.6 279 6 US-10-425-115-343835 Sequence 343835,  
43 74 9.6 358 5 US-09-248-796A-22578 Sequence 22578, A  
44 74 9.6 389 7 US-60-556-841-9640 Sequence 9640, Ap  
45 74 9.6 400 6 US-10-490-953-13 Sequence 13, Appl

#### ALIGNMENTS

##### RESULT 1

US-09-741-873C-4  
; Sequence 4, Application US/09741873C  
; GENERAL INFORMATION:  
; APPLICANT: Normark, Staffan  
; TITLE OF INVENTION: Fibronectin Binding Protein As Well As Its Preparation  
; FILE REFERENCE: 012889-084  
; CURRENT APPLICATION NUMBER: US/09/741,873C  
; CURRENT FILING DATE: 2000-12-22  
; PRIOR APPLICATION NUMBER: SE 8801723-1  
; PRIOR FILING DATE: 1998-05-06  
; PRIOR APPLICATION NUMBER: US 08/978,878  
; PRIOR FILING DATE: 1997-11-26  
; PRIOR APPLICATION NUMBER: US 07/347,189  
; PRIOR FILING DATE: 1989-05-04  
; PRIOR APPLICATION NUMBER: US 07/789,437  
; PRIOR FILING DATE: 1991-11-06  
; PRIOR APPLICATION NUMBER: US 07/970,846  
; PRIOR FILING DATE: 1992-11-03  
; PRIOR APPLICATION NUMBER: US 08/187,865  
; PRIOR FILING DATE: 1994-01-28  
; PRIOR APPLICATION NUMBER: US 08/318,519  
; NUMBER OF SEQ ID NOS: 11  
; SOFTWARE: Patentin version 3.0  
; SEQ ID NO 4  
; LENGTH: 151  
; TYPE: PRT  
; ORGANISM: Escherichia coli  
US-09-741-873C-4

Query Match 67.8%; Score 525; DB 5; Length 151;  
Best Local Similarity 68.2%; Pred. No. 9.3e-40;  
Matches 103; Conservative 21; Mismatches 27; Indels 0; Gaps 0;  
QY 1 MKLLKVAAPAAIVVSGSALAGVVPQWGGGNGHNGSSGPDSTLSIYQYGSANAALAQ 60  
Db 1 MKLLKVAAPAAIVVSGSALAGVVPQWGGGNGHNGSSGPDSTLSIYQYGSANAALAQ 60  
QY 61 SDARKYDQLVTRVVTHEMAHAGQAGADNSTIELTQNGFERNATTIDOWNAKNSDITVQYGG 120  
Db 61 TDARNSDLTITQHGNGGADVQGGSDSSIDLTRQFGNSATLDQNGKNSMTVKQFGG 120  
QY 121 NNAALVNCTASDSSVMVQVGFNNATANYQ 151  
Db 121 NGAAVDQTASNSSVNVTVQVGFNNATAHQY 151

US-09-741-873C-2  
; Sequence 2, Application US/09741873C  
; GENERAL INFORMATION:  
; APPLICANT: Normark, Staffan  
; APPLICANT: Olsen, Arne  
; TITLE OF INVENTION: Fibronectin Binding Protein As Well As Its Preparation  
; FILE REFERENCE: 012889-084  
; CURRENT APPLICATION NUMBER: US/09/741,873C  
; CURRENT FILING DATE: 2000-12-22  
; PRIOR APPLICATION NUMBER: SE 8801723-1  
; PRIOR FILING DATE: 1998-05-06  
; PRIOR APPLICATION NUMBER: US 08/978,878  
; PRIOR FILING DATE: 1997-11-26  
; PRIOR APPLICATION NUMBER: US 07/347,189  
; PRIOR FILING DATE: 1989-05-04  
; PRIOR APPLICATION NUMBER: US 07/789,437  
; PRIOR FILING DATE: 1991-11-06  
; PRIOR APPLICATION NUMBER: US 07/970,846  
; PRIOR FILING DATE: 1992-11-03  
; PRIOR APPLICATION NUMBER: US 08/187,865  
; PRIOR FILING DATE: 1994-01-28  
; PRIOR APPLICATION NUMBER: US 08/318,519  
; PRIOR FILING DATE: 1994-10-05  
; NUMBER OF SEQ ID NOS: 11  
; SOFTWARE: Patent in version 3.0  
; SEQ ID NO 2  
; LENGTH: 131  
; TYPE: PRT  
; ORGANISM: Escherichia coli  
US-09-741-873C-2

Query Match 57.8%; Score 447; DB 5; Length 131;  
Best Local Similarity 64.9%; Pred. No. 7,1e-33;  
Matches 88; Conservative 20; Mismatches 26; Indels 0; Gaps 0;

Qy 21 GVVPQWGGGNGHGGNSGPDSTLSIYQGSANAALQSDARKYDQLVTRVVTHEMAH 80  
Db 1 GVVPQWGGGNGHGGNSGPDSTLSIYQGSANAALQSDARKYDQLVTRVVTHEMAH 80

Qy 81 AGQADNSTIELTQNGFRNNATIDOWNAKNSDITVQYGGNNAALVNQTSASDSSVMVRQV 140  
Db 61 VQGSDDSSIDLQRFGENSATLDQNGKNSMTVTRVQFGGNGAAVDQTSASNVNTQV 120

Qy 141 GFGNNATANYQ 151  
Db 121 GFGNNATANYQ 131

RESULT 3  
US-60-565-632-7907  
; Sequence 7907, Application US/60565632  
; GENERAL INFORMATION:  
; APPLICANT: Monsanto Technology, LLC  
; APPLICANT: Baum, James A  
; APPLICANT: Kovalic, David K.  
; APPLICANT: Larosa, Thomas J.  
; APPLICANT: Lu, Maolong  
; APPLICANT: Munyikwa, Tichifa R. I.  
; APPLICANT: Roberts, James K.  
; APPLICANT: Wu, Wei  
; APPLICANT: Zhang, Bei  
; TITLE OF INVENTION: Methods for Genetic Control of Insect Infestations in Plants and  
; FILE REFERENCE: 38-21(53403)B  
; CURRENT APPLICATION NUMBER: US/60/565,632  
; CURRENT FILING DATE: 2004-04-27  
; NUMBER OF SEQ ID NOS: 15449  
; SOFTWARE: Patent in version 3.2  
; SEQ ID NO 7907  
; LENGTH: 573  
; TYPE: PRT  
; ORGANISM: Diabrotica virgifera

US-60-565-632-7907

Query Match 11.2%; Score 87; DB 7; Length 573;  
Best Local Similarity 24.9%; Pred. No. 9.5;  
Matches 42; Conservative 15; Mismatches 62; Indels 50; Gaps 7;

Qy 30 GNHNGGG--NSSGPDSTLSIYQGSANAALA-----LQSDARKYDQLVTRVVTHEMAH 80  
Db 241 GNENGTGAENNANADAQTDVAQ--GSTNEANAENNANADVQNDAAQANENGAAENSNGAD 299

Qy 81 AGQADN-STIELTON-----GFRNNATIDOWN----- 107  
Db 300 AAQGTDNGAAAEENTGNADPPAQGNDNGAAENSGNENGTAENNANADVQNDAAQVNDNGA 359

Qy 108 --AKNSDITVQYGGNNAALVNQTSAD-----SSVMVRQVQFGNNATAN 149  
Db 360 AENNGNADAQSDNDNGAAENNTNADAQNGAAQSTANEANAENNANAD 408

RESULT 4  
US-60-579-062-7907  
; Sequence 7907, Application US/60579062  
; GENERAL INFORMATION:  
; APPLICANT: Baum, James A  
; APPLICANT: Kovalic, David K  
; APPLICANT: Larosa, Thomas J  
; APPLICANT: Lu, Maolong  
; APPLICANT: Munyikwa, Tichifa R. I.  
; APPLICANT: Roberts, James K  
; APPLICANT: Wu, Wei  
; APPLICANT: Zhang, Bei  
; TITLE OF INVENTION: Methods for Genetic Control of Insect Infestations in Plants and  
; FILE REFERENCE: 38-21(53403)C  
; CURRENT APPLICATION NUMBER: US/60/579,062  
; CURRENT FILING DATE: 2004-06-11  
; NUMBER OF SEQ ID NOS: 41445  
; SOFTWARE: Patent in version 3.2  
; SEQ ID NO 7907  
; LENGTH: 573  
; TYPE: PRT  
; ORGANISM: Diabrotica virgifera  
US-60-579-062-7907

Query Match 11.2%; Score 87; DB 7; Length 573;  
Best Local Similarity 24.9%; Pred. No. 9.5;  
Matches 42; Conservative 15; Mismatches 62; Indels 50; Gaps 7;

Qy 30 GNHNGGG--NSSGPDSTLSIYQGSANAALA-----LQSDARKYDQLVTRVVTHEMAH 80  
Db 241 GNENGTGAENNANADAQTDVAQ--GSTNEANAENNANADVQNDAAQANENGAAENSNGAD 299

Qy 81 AGQADN-STIELTON-----GFRNNATIDOWN----- 107  
Db 300 AAQGTDNGAAAEENTGNADPPAQGNDNGAAENSGNENGTAENNANADVQNDAAQVNDNGA 359

Qy 108 --AKNSDITVQYGGNNAALVNQTSAD-----SSVMVRQVQFGNNATAN 149  
Db 360 AENNGNADAQSDNDNGAAENNTNADAQNGAAQSTANEANAENNANAD 408

RESULT 5  
US-10-425-115-93207  
; Sequence 193207, Application US/10425115  
; GENERAL INFORMATION:  
; APPLICANT: La Rosa, Thomas J.  
; APPLICANT: Kovalic, David K.  
; APPLICANT: Zhou, Yihua  
; APPLICANT: Cao, Yongwei  
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With  
; FILE REFERENCE: 38-21(53222)B  
; CURRENT APPLICATION NUMBER: US/10/425,115

```

; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 369326
; SEQ ID NO 193207
; LENGTH: 147
; TYPE: PRT
; ORGANISM: Zea mays
; FEATURE:
; OTHER INFORMATION: Clone ID: MRT4577_107790C.1.pep
US-10-425-115-193207

Query Match 11.0%; Score 85; DB 6; Length 147;
Best Local Similarity 24.6%; Pred. No. 2.8;
Matches 35; Conservative 18; Mismatches 63; Indels 26; Gaps 4;

QY 7 AAFAAVVGSSALAGVQPQGGGNGHGGSSGPDSTLSIYQY-----GSNAALALQSD 62
Db 27 AQIPASATTGAATAATPQPGSGTGGVVGAGGPDAPLELYMHDILGGSSPTA----- 80
QY 63 ARKYDQLVTRVTHEMAHAGGADNSTIELTQCFRNATIDQWAKNSDITVGYQGN 122
Db 81 -----RPIITGLLGNINYQVFPFAPFGFSAPRNGVAIPNANGQ-----VFTYNGT 126
QY 123 AALVNOTASDSSVMVRQVGFN 144
Db 127 GIPLDTGLSRAGFL--QPGTGN 146

RESULT 6
US-60-565-632-7905
; Sequence 7905, Application US/60565632
; GENERAL INFORMATION:
; APPLICANT: Monsanto Technology, LLC
; APPLICANT: Baum, James A
; APPLICANT: Kovalic, David K.
; APPLICANT: Larosa, Thomas J.
; APPLICANT: Lu, Maolong
; APPLICANT: Munyikwa, Tichifa R. I.
; APPLICANT: Roberts, James K.
; APPLICANT: Wu, Wei
; APPLICANT: Zhang, Bei
; TITLE OF INVENTION: Methods for Genetic Control of Insect Infestations in Plants and
; FILE REFERENCE: Compositions thereof
; CURRENT APPLICATION NUMBER: US/60/565.632
; CURRENT FILING DATE: 2004-04-27
; NUMBER OF SEQ ID NOS: 15449
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 7905
; LENGTH: 412
; TYPE: PRT
; ORGANISM: Diabrotica virgifera
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (369)..(369)
; OTHER INFORMATION: Xaa can be any naturally occurring amino acid
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (392)..(392)
; OTHER INFORMATION: Xaa can be any naturally occurring amino acid
US-60-565-632-7905

Query Match 11.0%; Score 85; DB 7; Length 412;
Best Local Similarity 26.4%; Pred. No. 9.6;
Matches 43; Conservative 12; Mismatches 62; Indels 46; Gaps 7;

QY 30 GNHNGGG--NSSGGPDSTLSIYQGSANAALA-----LQSDARKYDQLVTRVTHEMAH 80
Db 85 GNENGTGAENNANADAQTDAQ--GSTNEANENNANADVQNDAAAGENGAAENSGNAD 143
QY 81 AGQGADN--STIELTON-----GFRNATIDQW----- 107
Db 144 AAQCTDNGAAENTGNADPAQCNDNGAAENGNGTAAENNANADVQNDAAQVNNNGA 203

```

NUMBER OF SEQ ID NOS: 15449  
 SOFTWARE: Patentin version 3.2  
 SEQ ID NO 7906  
 LENGTH: 841

TYPE: PRT  
 ORGANISM: Diabrotica virgifera  
 NAME/KEY: misc feature  
 LOCATION: (810)..(810)  
 OTHER INFORMATION: Xaa can be any naturally occurring amino acid  
 US-60-565-632-7906

Query Match 11.0%; Score 85; DB 7; Length 841;  
 Best Local Similarity 30.7%; Pred. No. 23;  
 Matches 42; Conservative 13; Mismatches 56; Indels 26; Gaps 8;

QY 30 GNHGGG--NSSGPDSTLSIYQGSANAALA-----LQSDARKYDQVLRVVTHEMAH 80  
 DB 269 GNENGTGAENNAADACTDAQ--GSTNEAENNAADVQNDAAQANENGAAEENSGNAD 327  
 QY 81 AGQCADN-STIELTON-----GFRNNATIDQWNAKNSDITVGYGNNAAALVNOTASDSS 134  
 DB 328 AAQGTNDGAAEENTGNADPAQGNDCGAA-----AENSGNENGTAAGNNA---NPDVQND 379  
 QY 135 VMVRQVGF--NNATAN 149  
 DB 380 AQVNDNGTAENNGNAD 396

## RESULT 9

US-60-579-062-7906  
 Sequence 7906, Application US/60579062

GENERAL INFORMATION:  
 APPLICANT: Baum, James A  
 APPLICANT: Kovalic, David K  
 APPLICANT: Larosa, Thomas J  
 APPLICANT: Lu, Maolong  
 APPLICANT: Munyikwa, Tichifa R. I.  
 APPLICANT: Roberts, James K  
 APPLICANT: Zhang, Bei  
 TITLE OF INVENTION: Methods for Genetic Control of Insect Infestations in Plants and Compositions thereof  
 FILE REFERENCE: 38-21 (53403) C  
 CURRENT APPLICATION NUMBER: US/60/579,062  
 PRIOR FILING DATE: 2004-06-11  
 NUMBER OF SEQ ID NOS: 41445  
 SOFTWARE: Patentin version 3.2  
 SEQ ID NO 7906  
 LENGTH: 841  
 TYPE: PRT  
 ORGANISM: Diabrotica virgifera  
 NAME/KEY: misc feature  
 LOCATION: (810)..(810)  
 OTHER INFORMATION: Xaa can be any naturally occurring amino acid  
 US-60-579-062-7906

Query Match 11.0%; Score 85; DB 7; Length 841;  
 Best Local Similarity 30.7%; Pred. No. 23;  
 Matches 42; Conservative 13; Mismatches 56; Indels 26; Gaps 8;

QY 30 GNHGGG--NSSGPDSTLSIYQGSANAALA-----LQSDARKYDQVLRVVTHEMAH 80  
 DB 269 GNENGTGAENNAADACTDAQ--GSTNEAENNAADVQNDAAQANENGAAEENSGNAD 327  
 QY 81 AGQCADN-STIELTON-----GFRNNATIDQWNAKNSDITVGYGNNAAALVNOTASDSS 134  
 DB 328 AAQGTNDGAAEENTGNADPAQGNDCGAA-----AENSGNENGTAAGNNA---NPDVQND 379  
 QY 135 VMVRQVGF--NNATAN 149  
 DB 380 AQVNDNGTAENNGNAD 396

## RESULT 10

US-60-579-902-7377  
 Sequence 7377, Application US/60579902  
 GENERAL INFORMATION:  
 APPLICANT: Goldman, Barry S.  
 APPLICANT: Krasomil-Osterfeld, Karina C.  
 TITLE OF INVENTION: Nucleotide and Amino Acid Sequences from Xenorhabdus and Uses thereof  
 FILE REFERENCE: 38-21 (53444) A  
 CURRENT APPLICATION NUMBER: US/60/579,902  
 CURRENT FILING DATE: 2004-06-15  
 NUMBER OF SEQ ID NOS: 14985  
 SEQ ID NO 7377  
 LENGTH: 2663  
 TYPE: PRT  
 ORGANISM: Xenorhabdus bovienii  
 US-60-579-902-7377

Query Match 11.0%; Score 85; DB 7; Length 2663;  
 Best Local Similarity 24.7%; Pred. No. 91;  
 Matches 42; Conservative 23; Mismatches 59; Indels 46; Gaps 7;

QY 5 KVAFAFRAIVVSGSALAGVWPQGGGNGHNGSGPDSSTLSIYQGSANAALALQSDAR 64  
 DB 370 QLLASKSLVVSASAI-----DNFQKITSSGGDITLTGQLDNLGKIA---GQ 415  
 QY 65 KYDQVLRVVTHEMAHAGQAGDSTIELTQNGFRNA-----TIDWNAKNS- 111  
 DB 416 HHLQLNTQSINNEKGH--MSADTVDTNTHQGLNNTAGLIVAERNMLRTGELLNRQGSV 473  
 QY 112 ----DITVQYVG-----GNNAALVNOTASDSSVMVRQVGFNNATAN 149  
 DB 474 RSGQDLTLNTHGNQLDNRDSGLQGLFSQGA-----MHLNTGYNNQSGH 518

## RESULT 11

US-60-581-351-7636  
 Sequence 7636, Application US/60581351

GENERAL INFORMATION:  
 APPLICANT: Laurie, Cathy C  
 TITLE OF INVENTION: Nucleic Acid Molecules Associated With Protein In Plants  
 FILE REFERENCE: 38-21 (53372) B  
 CURRENT APPLICATION NUMBER: US/60/581,351  
 CURRENT FILING DATE: 2004-06-17  
 PRIOR APPLICATION NUMBER: US 60/479,962  
 PRIOR FILING DATE: 2003-06-19  
 NUMBER OF SEQ ID NOS: 13980  
 SOFTWARE: Patentin version 3.2  
 SEQ ID NO 7636  
 LENGTH: 436  
 TYPE: PRT  
 ORGANISM: Ceratopteris richardii  
 US-60-581-351-7636

Query Match 10.5%; Score 81.5; DB 7; Length 436;  
 Best Local Similarity 26.1%; Pred. No. 21;  
 Matches 29; Conservative 16; Mismatches 43; Indels 23; Gaps 3;

QY 29 GGNHNGGNS--SGPDSSTLSIYQGSANAALALQSDARKYDQVLRVVTHEMAHAGQGA 85  
 DB 77 GGGAGGGSSVAGGNTANTHNHQQHSAEAAGLLATARS-----GGDLAQSGQGG 127.  
 QY 86 DNSTIELTQNGFRNNATIDQWNAKNSDITVGYGNNAAALVNOTASDSSVM 136  
 DB 128 RGNLLDLHSD-----TANSSDLVEGEGHGGHGGHQQDSQVL 167

## RESULT 12

US-10-425-115-320950  
 Sequence 320950, Application US/10425115

```

; GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J.
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
; TITLE OF INVENTION: Plants
; FILE REFERENCE: 38-21(S3222)B
; CURRENT APPLICATION NUMBER: US/10/425.115
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 369326
; SEQ ID NO 320950
; LENGTH: 511
; TYPE: PRT
; ORGANISM: Zea mays
; FEATURE:
; NAME/KEY: unsure
; LOCATION: (1)..(511)
; OTHER INFORMATION: unsure at all Xaa locations
; FEATURE:
; OTHER INFORMATION: Clone ID: MRT4577_55769C.1.pep
; US_10-425-115-320950

Query Match      10.4%; Score 80.5; DB 6; Length 511;
Best Local Similarity 20.8%; Pred. No. 32;
Matches 38; Conservative 28; Mismatches 70; Indels 47; Gaps 6;

QY      1  MKLLKVAFAAIVWS-----GSAAGVVPQWGG-----GNHNGGNSSGPD 42
DB      133  LALLSERAQAILASCRDVLASGFRFEDAWKVIQSGDGVYVAWVAANYALGRLGDDN 192

QY      43  STLSIYQVGSANAUALOSDA-----RKYDQLVTRVVTTHMAHAGGADNSTIE-- 91
DB      193  KITVGIIELGSAQLTFVSDEVLPPKLSYNTFGETTYTLYTNSFLNFGNAQAQDSFHEM 252

QY      92  -LTONGFRNNATID---CWAKSDDITVQYGNNALVNOTASDSVMVVRQVGNNNA 146
DB      253  LKRSQFKNGTLADCAPRGYSRNEEMLRMSGASRSTLENQ-----YVNNG 299

QY      147  TAN 149
DB      300  TGN 302

```

```

RESULT 13
PCT-US04-07412-1695
/ Sequence 1695, Application PC/TUS0407412
/ GENERAL INFORMATION:
/ APPLICANT: Tang, Y. Tom
/ APPLICANT: Wang, Jian-rui
/ APPLICANT: Zhang, Jie
/ APPLICANT: Ren, Feiyan
/ APPLICANT: Zhou, Ping
/ APPLICANT: Ma, Yunding
/ APPLICANT: Ghosh, Malabika
/ APPLICANT: Xue, Aidong J.
/ APPLICANT: Aesundi, Vinod
/ APPLICANT: Zhao, Qung A.
/ APPLICANT: Wang, Dnruai
/ APPLICANT: Goodrich, Ryle W.
/ APPLICANT: Chen, Rui-hong
/ APPLICANT: Wehrman, Tom
/ APPLICANT: Wang, Gezhi
/ APPLICANT: Wang, Zhiwei
/ APPLICANT: Boyle, Bryan J.
/ APPLICANT: Drmanac, Radoje T.
/ TITLE OF INVENTION: Novel Nucleic Acids
/ FILE REFERENCES: 822CIP/PCT
/ CURRENT APPLICATION NUMBER: PCT/US04/07412
/ PRIOR FILING DATE: 2004-03-19
/ PRIOR APPLICATION NUMBER: US 10/389,559
/ PRIOR FILING DATE: 2003-03-14
/ PRIOR APPLICATION NUMBER: US 60/365,264

```

[illegible]

RESULT 14  
US-10-389-559-1695  
Sequence 1695, Application US/10389559  
GENERAL INFORMATION:  
APPLICANT: Tang, Y. Tom  
APPLICANT: Wang, Jian-rui  
APPLICANT: Zhang, Jie  
APPLICANT: Ren, Feiyan  
APPLICANT: Zhou, Ping  
APPLICANT: Ma, Yuning  
APPLICANT: Ghosh, Malabika  
APPLICANT: Xue, Aidong J.  
APPLICANT: Asundi, Vinod  
APPLICANT: Zhao, Qing A.  
APPLICANT: Wang, Dunrui  
APPLICANT: Goodrich, Ryle W.  
APPLICANT: Chen, Rui-hong  
APPLICANT: Wehrman, Tom  
APPLICANT: Wang, Gezhi  
APPLICANT: Wang, Zhiwei  
APPLICANT: Boyle, Bryan J.  
APPLICANT: Drmanac, Radoje T.  
TITLE OF INVENTION: Novel Nucleic Acids and Polypeptides  
FILE REFERENCE: 822  
CURRENT APPLICATION NUMBER: US/10/389,559  
CURRENT FILING DATE: 2003-03-14  
PRIOR APPLICATION NUMBER: US 60/365,254  
PRIOR FILING DATE: 2002-03-14  
PRIOR APPLICATION NUMBER: US 60/340,187  
PRIOR FILING DATE: 2001-12-12

us-09-543-407-20.rapn

Wed Aug 4 10:14:27 2004

OY 128 QTASDSSVMVQVGF 143  
 Db 276 LDKAKSCFAMGQAGG 291  
 Search completed: August 2, 2004, 15:29:53  
 Job time : 18.8 secs

PRIOR APPLICATION NUMBER: PCT/US00/35017  
 PRIOR FILING DATE: 2000-12-22  
 PRIOR APPLICATION NUMBER: US 09/552,317  
 PRIOR FILING DATE: 2000-04-25  
 PRIOR APPLICATION NUMBER: US 09/488,725  
 PRIOR FILING DATE: 2000-01-21  
 PRIOR APPLICATION NUMBER: PCT/US01/02623  
 PRIOR FILING DATE: 2001-01-25  
 PRIOR APPLICATION NUMBER: US 09/491,404  
 PRIOR FILING DATE: 2000-01-25  
 PRIOR APPLICATION NUMBER: PCT/US01/03800  
 PRIOR FILING DATE: 2001-02-05  
 PRIOR APPLICATION NUMBER: US 09/560,875  
 PRIOR FILING DATE: 2000-04-27  
 PRIOR APPLICATION NUMBER: US 09/496,914  
 PRIOR FILING DATE: 2000-02-03  
 Remaining Prior Application data removed - See File Wrapper or PALM.  
 NUMBER OF SEQ ID NOS: 1920  
 SOFTWARE: PCT-FL Genes Version 6.0  
 SEQ ID NO 1695  
 LENGTH: 581  
 TYPE: PRT  
 ORGANISM: Homo sapiens  
 US-10-369-559-1695

Query Match 10.3%; Score 80; DB 6; Length 581;  
 Best Local Similarity 22.3%; Pred. No. 41;  
 Matches 27; Conservative 25; Mismatches 47; Indels 22; Gaps 5;

OY 45 LSIYQY-GGANAALQSDARKYDQ--VTRVVTHEMAHAGQADNSTIELTQNGFRNNA 101  
 Db 242 VSLYPLAGGTQAFKENNQKAYETVGVSHITRDMLQIPKQQQNEKYQVPQ---FDQS 298  
 OY 102 TI-----DOWNAKNSDITVQYGGNNAALV----NQFASDSSVMVQVGFNN 145  
 Db 299 TIKNIESAKGLDVMDSWPLQNDGTVAEYGVHVPFALAGSPKADDTTSIYFYOKVGDN 358  
 OY 146 A 146  
 Db 359 S 359

RESULT 15  
 US-60-556-841-11319  
 Sequence 11319, Application US/60556841  
 GENERAL INFORMATION:  
 APPLICANT: Abad, Mark S.  
 TITLE OF INVENTION: Genes and Uses for Plant Improvement  
 FILE REFERENCE: 38-21(53450)  
 CURRENT APPLICATION NUMBER: US/60/556,841  
 CURRENT FILING DATE: 2004-03-25  
 NUMBER OF SEQ ID NOS: 12463  
 SEQ ID NO 11319  
 LENGTH: 321  
 TYPE: PRT  
 ORGANISM: Caulobacter crescentus CB15  
 US-60-556-841-11319

Query Match 10.3%; Score 79.5; DB 7; Length 321;  
 Best Local Similarity 24.5%; Pred. No. 22;  
 Matches 48; Conservative 15; Mismatches 64; Indels 69; Gaps 10;

OY 4 LKVAFAAIVVSGA-----LAG---VVPQGGGGNHG----- 34  
 Db 109 LKAAGLKMVMVCGAVKHAKAEQACDAVICQGGEGGHTGLVGTPLVAQAVEAVKIPV 168  
 OY 35 ---CGNSGSPDSTLSIYQGSANAALQSDARKYDOLVTRVVTHEMAHAG-----Q 83  
 Db 169 VAAGGLHDG-----RGLA-LALALGAGQ---VVMGTFRFIASHEAHAGDLYRQAVVE 215  
 OY 84 GADNSTIEL-TQNG-----FRNATIDOWNAKNSDITV-----GQYGGNNAALVN 127  
 Db 216 AADEDTVTRCYSGKPMRVKKNPYDDWEARPGDIQFPQOAMVSRINGAMGGIGGQIEG 275



GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: August 2, 2004, 14:39:53 ; Search time 9.4 Seconds

(without alignments)  
1545.204 Million cell updates/sec

Title: US-09-543-407-20

Perfect score: 774  
Sequence: 1 MKLLKVAFAALVVGSSALA.....DSSVMVRQVGFNNATANQY 151

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 283366 seqs, 96191526 residues

Total number of hits satisfying chosen parameters: 283366

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : PIR 78:\*

1: PIR1:\*

2: PIR2:\*

3: PIR3:\*

4: PIR4:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	689	89.0	151	2 JC6039	fimbrin protein ag
2	689	89.0	151	2 A10635	major curlin chain
3	528	68.2	151	2 S70788	curlin protein csg
4	506.5	65.4	152	2 D90806	curlin major subun
5	506.5	65.4	152	2 H85665	hypothetical prote
6	105.5	13.6	145	2 AD3143	conserved hypothet
7	105.5	13.6	145	2 H98144	hypothetical prote
8	99	12.8	573	2 C86266	F319.21 protein -
9	96.5	12.5	151	2 S70787	curlin nucleator p
10	96.5	12.5	151	2 C90806	minor curlin subun
11	96.5	12.5	151	2 G85665	curlin minor chain
12	96	12.4	2174	2 E95965	hypothetical glyci
13	92.5	12.0	1567	2 S11672	ice nucleation pro
14	91.5	11.8	1322	2 S07053	ice nucleation pro
15	91	11.8	602	1 PL0221	leishmanolysin (EC
16	91	11.8	645	2 F70825	probable PPE prote
17	90	11.6	639	2 C42049	leishmanolysin (EC
18	89.5	11.6	374	2 T03875	probable homeobox
19	89.5	11.6	375	2 T03874	probable homeobox
20	89.5	11.6	447	2 G84687	probable disease r
21	89	11.5	652	2 E97857	cell surface antig
22	88.5	11.4	151	2 JC6040	fimbrin protein ag
23	88.5	11.4	151	2 AH0635	nucleation compone
24	88.5	11.4	1588	2 A86036	probable adhesin Z
25	88.5	11.4	1588	2 H91188	probable adhesin E
26	88	11.4	590	1 A45621	leishmanolysin (EC
27	88	11.4	599	2 B42049	leishmanolysin (EC
28	88	11.4	599	2 A44951	leishmanolysin (EC
29	88	11.4	646	1 S19916	leishmanolysin (EC

#### ALIGNMENTS

##### RESULT 1

JC6039

fimbrin protein agFA precursor - Salmonella enteritidis

C:Species: Salmonella enteritidis

C:Date: 31-Dec-1996 #sequence\_revision 31-Dec-1996 #text\_change 08-Oct-1999

R:Accession: JC6039; PC6015; A44898

R:Collinson, S.K.; Clouthier, S.C.; Doran, J.L.; Baner, P.A.; Kay, W.W.

J. Bacteriol. 178, 662-667, 1996

A:Title: Salmonella enteritidis agfBAC operon encoding thin, aggregative fimbriae.

A:Reference number: JC6039; MUID:96146512; PMID:8550497

A:Accession: JC6039

A:Molecule type: DNA

A:Residues: 1-151 <COL>

A:Cross-references: GB:U43280; NID:g1184712; PIDN:AAC43599.1; PID:g1184714

A:Accession: PC6015

A:Molecule type: protein

A:Residues: 21-52 <CO2>

A:Experimental source: strain 27655-3b

A:Note: the authors translated the codon ACG for residue 44 as Ile

R:Collinson, S.K.; Emody, L.; Muller, K.H.; Trust, T.J.; Kay, W.W.

J. Bacteriol. 173, 4773-4781, 1991

A:Title: Purification and characterization of thin, aggregative fimbriae from Salmonella

A:Reference number: A44898; MUID:91310586; PMID:1677357

A:Contents: 27655

A:Accession: A44898

A>Status: preliminary

A:Molecule type: protein

A:Residues: 21-33 <CO3>

A:Note: sequence extracted from NCBI backbone (NCBIP:45936)

C:Genetics:

A:Gene: agFA

C:Function:

A:Description: major component of thin aggregative fimbriae

A:Note: fimbriae bind to fibronectin, plasminogen, tissue plasminogen activator

C:Keywords: fimbria

F:1-20/Domain: signal sequence #status predicted <SIG>

F:21-151/Product: fimbrin protein agFA #status experimental <MAT>

Query Match 89.0%; Score 689; DB 2; Length 151;

Best Local Similarity 90.7%; Pred.No. 3.1e-51;

Matches 137; Conservative 3; Mismatches 11; Indels 0; Gaps 0;

Qy 1 MKLLKVAFAALVVGSSALAGVVPQWGGGNGHNGGNSGPDSTLSIYQYGSANAALQ 60

|||||

Db 1 MKLLKVAFAALVVGSSALAGVVPQWGGGNGHNGGNSGPDSTLSIYQYGSANAALQ 60

|||||

Qy 61 SDARKYQDLVTRVTHEMAHAGQADNSTIELTQNGFRNNATIDOWNAKNSDITVGYGG 120

|||||

Db 61 SDARKSETTITQSGYNGADVGQGDNSTIELTQNGFRNNATIDOWNAKNSDITVGYGG 120

|||||

Qy 121 NNAALVNOTASDSSVMVRQVGFNNATANQY 151

|||||

Db 121 NNAALVNQTASDSSVMVRQVGFNNATANQY 151

RESULT 2

AI0635

major curlin chain precursor [imported] - Salmonella enterica subsp. enterica serovar Typhimurium

C:Species: Salmonella enterica subsp. enterica serovar Typhi

A:Note: This species has also been called Salmonella typhi

C>Date: 09-Nov-2001 #sequence\_revision 09-Nov-2001 #text\_change 18-Nov-2002

C:Accession: AI0635

R:Parkhill, J.; Dougan, G.; James, K.D.; Thomson, N.R.; Pickard, D.; Wain, J.; Churcher, T.; Connor, P.; Cronin, A.; Davis, P.; Davies, R.M.; Dowd, L.; White, N.; Farrar, S.; Moule, S.; O'Gaora, P.

Nature 413, 848-852, 2001

A:Authors: Parry, C.; Quail, M.; Rutherford, K.; Simmonds, M.; Skelton, J.; Stevens, K.

A:Title: Complete genome sequence of a multiple drug resistant Salmonella enterica serovar Typhimurium

A:Reference number: AB0502; MUID:21534947; PMID:11677608

A:Accession: AI0635

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-151 <PAR>

A:Cross-references: GB:AL513382; PIDN:CAD08268.1; PID:G16502315; GSPDB:GN00176

C:Genetics: S11181

A:Gene: S11181

Query Match 89.0%; Score 689; DB 2; Length 151;

Best Local Similarity 90.7%; Pred. No. 3.1e-51;

Matches 137; Conservative 3; Mismatches 11; Indels 0; Gaps 0;

QY 1 MKLLKVAFAAIVVSGSALAGVVPQGGGNGHNGGNSGPDSTLSIYQVGSANAALALQ 60

DB 1 MKLLKVAFAAIVVSGSALAGVVPQGGGNGHNGGNSGPDSTLSIYQVGSANAALALQ 60

QY 61 SDARKYDQLVTRVVTHEMAHAGGADNSTIETQNGFRNNATIDOWNAKNSDITVQYGG 120

DB 61 SDARKSETTITQSGYNGADVGAGDNSTIETQNGFRNNATIDOWNAKNSDITVQYGG 120

QY 121 NNAALVNQTASDSSVMVRQVGFNNATANQY 151

DB 121 NNAALVNQTASDSSVMVRQVGFNNATANQY 151

RESULT 3

S70788

curlin protein csgA precursor - Escherichia coli (strain K-12)

N:Alternate names: csgA protein; major curlin protein

C:Species: Escherichia coli

C>Date: 12-Feb-1998 #sequence\_revision 20-Feb-1998 #text\_change 01-Mar-2002

C:Accession: S70788; G64846; S31202; S34560; S34559

R:Hammar, M.; Arngqvist, A.; Bian, Z.; Olsen, A.; Normark, S.

Mol. Microbiol. 18, 661-670, 1995

A:Title: Expression of two csg operons is required for production of fibronectin- and curli

A:Reference number: S70788; MUID:96414468; PMID:8817489

A:Accession: S70788

A:Status: nucleic acid sequence not shown; translation not shown

A:Molecule type: DNA

A:Residues: 1-151 <HAM>

A:Cross-references: EMBL:X90754; NID:g1147558; PIDN:CAA62282.1; PID:g1147564

A:Experimental source: strain K12, substrain W3110

A:Note: The nucleotide sequence was submitted to the EMBL Data Library, August 1995

R:Blattner, F.R.; Plunkett III, G.; Bloch, C.A.; Ferna, N.T.; Burland, V.; Riley, M.; Cohen, D.; Rose, D.J.; Mau, B.; Shao, Y.

Science 277, 1453-1462, 1997

A:Title: The complete genome sequence of Escherichia coli K-12.

A:Reference number: A64720; MUID:97426617; PMID:9278503

A:Accession: G64846

A:Status: nucleic acid sequence not shown; translation not shown

A:Molecule type: DNA

A:Residues: 1-151 <BLAT>

A:Cross-references: GB:AE000205; GB:U00096; NID:g1787265; PIDN:AACT4126.1; PID:g1787279;

A:Experimental source: strain K-12, substrain MG1655

R:Olsen, A.; Arngqvist, A.; Hammar, M.; Sukupolvi, S.; Normark, S.

Mol. Microbiol. 7, 523-536, 1993

A:Title: The RpoS sigma factor relieves H-NS-mediated transcriptional repression of csgA.

A:Reference number: S31202; MUID:93211294; PMID:8459772

A:Accession: S31202

A:Molecule type: DNA

A:Residues: 1-6,'V', 8-151 <OLS1>

A:Cross-references: EMBL:L04979

A:Accession: S34560

A:Molecule type: protein

A:Residues: 21-42;44-50 <OLS2>

R:Olsen, A.N.; Arngqvist, A.M.

submitted to the EMBL Data Library, October 1992

A:Reference number: S34559

A:Accession: S34559

A:Molecule type: DNA

A:Residues: 1-133,'RQRSGWLV' <OLS3>

A:Cross-references: EMBL:L04979; NID:g290424; PIDN:AAA23616.1; PID:g290425

A:Experimental source: strain K-12, substrain W3110

C:Genetics:

A:Gene: csgA

A:Map position: 23.15

C:Function:

A:Description: major component of wild-type curli; interaction between CsgA and CsgB triggers curli assembly

A:Note: curli are thin, coiled fibers expressed on the surface of Escherichia coli that bind to H-kininogen; in the absence of CsgA, CsgB can self-assemble into polymers

F:1-20/Domain: signal sequence #status predicted <SIG>

P:21-151/Product: curlin #status experimental <MAT>

Query Match 68.2%; Score 528; DB 2; Length 151;

Best Local Similarity 68.9%; Pred. No. 1.2e-37;

Matches 104; Conservative 20; Mismatches 27; Indels 0; Gaps 0;

QY 1 MKLLKVAFAAIVVSGSALAGVVPQGGGNGHNGGNSGPDSTLSIYQVGSANAALALQ 60

DB 1 MKLLKVAFAAIVVSGSALAGVVPQGGGNGHNGGNSGPDSTLSIYQVGSANAALALQ 60

QY 61 SDARKYDQLVTRVVTHEMAHAGGADNSTIETQNGFRNNATIDOWNAKNSDITVQYGG 120

DB 61 TDARNSDLTITQHGCGNGADVGQGGSDSSIDLTRQFGNSATLQDNGKNSMTVQYFGG 120

QY 121 NNAALVNQTASDSSVMVRQVGFNNATANQY 151

DB 121 GNGAIVDTASNSVNTVQVGFNNATANQY 151

RESULT 4

D90806

curlin major subunit CsgA [imported] - Escherichia coli (strain O157:H7, substrain RIMD 0509952)

C:Species: Escherichia coli

C>Date: 18-Jul-2001 #sequence\_revision 18-Jul-2001 #text\_change 18-Jul-2001

C:Accession: D90806

R:Hayashi, T.; Makino, K.; Kurokawa, K.; Ishii, K.; Yokoyama, K.; Han, C.G.; Gasawara, N.; Yasunaga, T.; Kuhara, S.; Shiba, T.; Hattori, M.; Shinagawa, H.

DNA Res. 8, 11-22, 2001

A:Title: Complete genome sequence of enterohemorrhagic Escherichia coli O157:H7 and genomic islands

A:Reference number: A99629; MUID:21156231; PMID:11258796

A:Accession: D90806

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-152 <HAY>

A:Cross-references: GB:BA000007; PIDN:BA034843.1; PID:g13360880; GSPDB:GN00154

A:Experimental source: strain O157:H7, substrain RIMD 0509952

C:Genetics:

A:Gene: Ecs1420

Query Match 65.4%; Score 506.5; DB 2; Length 152;

Best Local Similarity 67.1%; Pred. No. 7.7e-36;

Matches 102; Conservative 21; Mismatches 28; Indels 1; Gaps 1;

QY 1 MKLLKVAFAAIVVSGSALAGVVPQGGGNGHNGGNSGPDSTLSIYQVGSANAALAL 59

DB 1 MKLLKVAFAAIVVSGSALAGVVPQGGGNGHNGGNSGPDSTLSIYQVGSANAALAL 60

QY 60 QSDARKYDQLVTRVVTHEMAHAGGADNSTIETQNGFRNNATIDOWNAKNSDITVQYGG 119

```
Db 61 QADARNSDLTITOHGGGAGADVQGGSDSSIDITQRFNGSATLDOWNGKDSHMTVKQFG 120
QY 120 GNAALVNQTSVSSVMVQVGFNNATANQY 151
Db 121 GNGAAVDQTASNSTVNTVQVGFNNATAHQY 152

RESULT 5
H85665
hypothetical protein csgA [imported] - Escherichia coli (strain O157:H7, substrain EDL933)
C:Species: Escherichia coli
C>Date: 16-Feb-2001 #sequence_revision 16-Feb-2001 #text_change 14-Sep-2001
C:Accession: H85665
R:Perna, N.T.; Plunkett III, G.; Burland, V.; Mau, B.; Glasner, J.D.; Rose, D.J.; Mayhew,
Miller, L.; Grotbeck, E.J.; Davis, N.M.; Lim, A.; Dimalanta, E.; Potamousis, K.; Apodaca,
Nature 408, 529-533, 2001
A:Title: Genome sequence of enterohemorrhagic Escherichia coli O157:H7.
A:Reference number: A85480; MUID:21074935; PMID:11206551
A:Accession: H85665
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-152 <STO>
A:Cross-references: GB:AE005174; NID:gl2514574; PIDN:AA655788.1; GSPDB:GN00145; UMGF:Z18
A:Experimental source: strain O157:H7, substrain EDL933
C:Genetics:
A:Gene: csgA

Query Match 65.4%; Score 506.5; DB 2; Length 152;
Best Local Similarity 67.1%; Pred. No. 7.7e-36;
Matches 102; Conservative 21; Mismatches 28; Indels 1; Gaps 1;

QY 1 MKLLKVAFAAIVVSGSALAGVVPQW-GGGNGHNGGSSGPDSTLSIYQYGSANAALAL 59
Db 1 MKLLKVAATAAIVFSGSALAGVVPQYGGGGHGGGNSGENSELNIYQYGGNSALAL 60

QY 60 QSDARKYDQVLTVRVTHEMAHAGGADNSTIELTQNGFRNNATIDOWNAKNSDITVQGVG 119
Db 61 QADARNSDLTITOHGGGAGADVQGGSDSSIDITQRFNGSATLDOWNGKDSHMTVKQFG 120

QY 120 GNAALVNQTSVSSVMVQVGFNNATANQY 151
Db 121 GNGAAVDQTASNSTVNTVQVGFNNATAHQY 152

RESULT 6
AD3143
conserved hypothetical protein Atu4768 [imported] - Agrobacterium tumefaciens (strain C58)
C:Species: Agrobacterium tumefaciens
C>Date: 11-Jan-2002 #sequence_revision 11-Jan-2002 #text_change 18-Nov-2002
C:Accession: AD3143
R:Wood, D.W.; Setubal, J.C.; Kaul, R.; Monks, D.; Chen, L.; Wood, G.E.; Chen, Y.; Woo, I.;
erage, G.; Gillet, W.; Grant, C.; Guenther, D.; Kutayavin, T.; Levy, R.; Li, M.; McClell
; Karp, P.; Romero, P.; Zhang, S.
Science 294, 2317-2323, 2001
A:Authors: Yoo, H.; Tao, Y.; Biddle, P.; Jung, M.; Krespan, W.; Perry, M.; Gordon-Kamm,
ster, E.W.
A:Title: The Genome of the Natural Genetic Engineer Agrobacterium tumefaciens C58.
A:Reference number: AB2577; MUID:21608550; PMID:11743193
A:Accession: AD3143
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-145 <KUR>
A:Cross-references: GB:AE008689; PIDN:AA45562.1; PID:gl17743277; GSPDB:GN00187
A:Experimental source: strain C58 (Dupont)
C:Genetics:
A:Gene: Atu4768
A:Map position: linear chromosome

Query Match 13.6%; Score 105.5; DB 2; Length 145;
Best Local Similarity 23.2%; Pred. No. 0.049;
Matches 36; Conservative 27; Mismatches 61; Indels 31; Gaps 4;

Db 61 QADARNSDLTITOHGGGAGADVQGGSDSSIDITQRFNGSATLDOWNGKDSHMTVKQFG 120
QY 120 GNAALVNQTSVSSVMVQVGFNNATANQY 151
Db 121 GNGAAVDQTASNSTVNTVQVGFNNATAHQY 152

RESULT 7
H98144
hypothetical protein AGR_L_228 [imported] - Agrobacterium tumefaciens (strain C58, Cereor
C:Species: Agrobacterium tumefaciens
C>Date: 22-Oct-2001 #sequence_revision 22-Oct-2001 #text_change 18-Nov-2002
C:Accession: H98144
R:Goodner, B.; Wollam, C.; Allinger, M.; Doughty, D.; Scott, C.; Lappas, C.; Markelz, B.;
A.; Liu, F.; Wollam, C.; Allinger, M.; Doughty, D.; Scott, C.; Lappas, C.; Markelz, B.;
Science 294, 2323-2328, 2001
A:Title: Genome Sequence of the Plant Pathogen and Biotechnology Agent Agrobacterium tum
A:Reference number: A97359; MUID:21608551; PMID:11743194
A:Accession: H98144
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-145 <KUR>
A:Cross-references: GB:AE007870; PIDN:AAK88692.1; PID:gl5158413; GSPDB:GN00170
C:Genetics:
A:Gene: AGR_L_228
A:Map position: linear chromosome

Query Match 13.6%; Score 105.5; DB 2; Length 145;
Best Local Similarity 23.2%; Pred. No. 0.049;
Matches 36; Conservative 27; Mismatches 61; Indels 31; Gaps 4;

QY 3 LLKVAFAAIVVSGSALAGVVPQW-GGGNGHNGGSSGPDSTLSIYQYGSANA 55
Db 1 MIRKSFTASALVALVGLSAAAPAMANDVRIBOYGSNSAGGAQEGYGNRIITYQNGGYNR 60

QY 56 ALALQSDARKYDQVLTVRVTHEMAHAGGADNSTIELTQNGFRNNATIDOWNAKNSDITV 115
Db 61 IVGHQ-----YGR-----HNL SAVQGEHDNYGTSITQNGNRNVAGI----- 96

QY 116 QYGGNNAALVNQTSVSSVMVQVGFNNATANQ 150
Db 97 GQFGSNHTTITQDNGNGNIAAGVQVGRGCSANVSQ 131

RESULT 8
C86266
F3F19.21 protein - Arabidopsis thaliana
C:Species: Arabidopsis thaliana (mouse-ear cress)
C>Date: 02-Mar-2001 #sequence_revision 02-Mar-2001 #text_change 09-Nov-2001
C:Accession: C86266
R:Theologis, A.; Ecker, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White, O.; Alonso,
Chen, C.W.; Chung, M.K.; Korn, L.; Conway, A.B.; Conway, A.R.; Creasy, T.H.; Dewar, K.;
ansen, N.F.; Hughes, B.; Huizar, L.
Nature 408, 816-820, 2000
A:Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khavkin, E.; Kim, C.;
C.A.; Li, J.H.; Li, Y.; Lin, X.; Liu, S.X.; Liu, Z.A.; Luros, J.S.; Maiti, R.; Marziali,
Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.
A:Authors: Salzberg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.; Tallon, I.
ker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.
A:Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.
A:Reference number: A86141; MUID:21016719; PMID:11130712
A:Accession: C86266
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-573 <STO>
A:Cross-references: GB:AE005172; NID:94850402; PIDN:AA031072.1; GSPDB:GN00141
C:Genetics:
```

```
Query Match 13.6%; Score 105.5; DB 2; Length 145;
Best Local Similarity 23.2%; Pred. No. 0.049;
Matches 36; Conservative 27; Mismatches 61; Indels 31; Gaps 4;
```



A;Reference number: A95842; MUID:21396508; PMID:11481431

A;Accession: E95965

A;Status: preliminary

A;Molecule type: DNA

A;Residues: 1-2174 <KUR>

A;Cross-references: GB:AL591985; PIDN:CAK49389.1; PID:g15140875; GSPDB:GN00167

A;Experimental source: strain 1021, megaplasmid pSymb

R;Galibert, F.; Finan, T.M.; Long, S.R.; Puhler, A.; Abola, P.; Ampe, F.; Barloy-Hubler,

Pella, D.; Chain, P.; Cowie, A.; Davis, R.W.; Dreano, S.; Federspiel, N.A.; Fisher, R.F.;

L.; Hyman, R.W.; Jones, T.

Science 293, 668-672, 2001

A;Authors: Kahn, D.; Kahn, M.L.; Kalman, S.; Keating, D.H.; Kiss, E.; Komp, C.; Lelaure,

hehault, P.; Vandenbol, M.; Vorholter, F.J.; Weidner, S.; Wells, D.H.; Wong, K.; Yeh, K.

A;Title: The composite genome of the legume symbiont *Sinorhizobium meliloti*.

A;Reference number: A96039; MUID:21368234; PMID:11474104

A;Contents: annotation

C;Genetics:

A;Gene: SMD21548

A;Genome: plasmid

Query Match 12.4%; Score 96; DB 2; Length 2174;  
Best Local Similarity 27.0%; Pred. No. 6.5;  
Matches 40; Conservative 18; Mismatches 54; Indels 36; Gaps 7;

QY 11 AIVVSGSALAGVVPQ--WGGGNGHNGGSSGPDSTLSIYQVGS-----ANAA----- 56

DB 693 AIATAGAGAVGILAQSIGGGGN--GGNATGGDAGFGFQIGGGGGGGYANTANVGFK 749

QY 57 -LALQSDARKYDQLVTRVVTHEMAHAGQADNSTIELTQNGFRNNATIDQWNAKNSDITV 115

DB 750 GLTLTTQGSAAAGIVAQSV-----GGGGTGTASSYAGIGFTAS-----VAV 793

QY 116 GOYGGNAA--LVNQTASDSSVMVRQV 141

DB 794 GGTGGNGAGGEVSLSITSAIRTGGQ 821

#### RESULT 13

S11672

ice nucleation protein - Xanthomonas campestris

C;Species: Xanthomonas campestris

C;Date: 21-Nov-1993 #sequence\_revision 10-Nov-1995 #text\_change 26-Aug-1999

A;Accession: S11672

A;Status: preliminary

A;Molecule type: DNA

A;Residues: 1-1567 <ZHA>

A;Cross-references: EMBL:X52970; NID:948531; PIDN:CAA37140.1; PID:g48532

C;Superfamily: ice nucleation protein

Query Match 12.0%; Score 92.5; DB 2; Length 1567;

Best Local Similarity 25.8%; Pred. No. 8.9;

Matches 46; Conservative 27; Mismatches 56; Indels 49; Gaps 11;

QY 14 VSSGALAG-----VVFQWGG--GGNHN-----GGNNSGPDSTLSIYQVGSANNAAL 59

DB 205 VYGTLTGADQSLVAGYGTETAGDHSLLIAGYGTGTAGSDSSI-LAGYGSTQTAAGR 263

QY 60 QSDARKYDQLVLT-----RVVTHEMAHAGQADNSTIELTQNGFRNNATI----- 103

DB 264 STLTAGYGSTQTAQEGSRLTSYGVSTATSGSDSAVI-----SGYGSTQTAGSSSLTAGYG 319

QY 104 -DOWNAKNSDITVQVYG-----GNNALV-----NOTASDSSVMVRQVFGNNATANQ 150

DB 320 STQTARKGSDITAG-YGSTGTAGSDSALIAGYGSTQTAGSESSLT--AGYGSTQTARK 374

#### RESULT 14

S07053

ice nucleation protein inaa - Erwinia ananas

C;Species: Erwinia ananas

C;Date: 30-Sep-1991 #sequence\_revision 30-Sep-1991 #text\_change 26-Aug-1999

C;Accession: S07053

R;Abe, K.; Watabe, S.; Emori, Y.; Watanabe, M.; Arai, S.

FEBS Lett. 258, 297-300, 1989

A;Title: An ice nucleation active gene of *Erwinia ananas*. Sequence similarity to those of

A;Reference number: S07053; MUID:90092494; PMID:259095

A;Accession: S07053

A;Status: not compared with conceptual translation

A;Molecule type: DNA

A;Residues: 1-1322 <AB>

A;Cross-references: GB:X17316; NID:9296095; PIDN:CAA35194.1; PID:g296096

C;Superfamily: ice nucleation protein

Query Match 11.8%; Score 91.5; DB 2; Length 1322;  
Best Local Similarity 29.1%; Pred. No. 8.9;  
Matches 37; Conservative 19; Mismatches 22; Indels 49; Gaps 9;

QY 34 GGGNSGPDSTLSIYQVGSANNAALQSDARKYDQLVTRVVTHEMAHAGQADNSTIELT 93

DB 933 GSTTAGPDSSL-AGYGSTQTA-----GYSILT-----AGYGS-----T 967

QY 94 QNGFRNNATIDQWNAKNSDITVGOYG-----GNNALV-----NOTASDSSVMVRQVFG 143

DB 968 QTG-----QENSDLITG-YGSTTAGYESSLIAGYGSTQTASFKSLTM--AGYG 1013

QY 144 NNATANQ 150

DB 1014 SSQTARE 1020

#### RESULT 15

PL0221

leishmanolysin (EC 3.4.24.36) precursor [validated] - Leishmania major

N;Alternate names: promastigote surface proteinase; surface endopeptidase glycoprotein gp

C;Species: Leishmania major

C;Date: 16-Sep-1992 #sequence\_revision 16-Sep-1992 #text\_change 15-Sep-2000

C;Accession: PL0221; A27598; A60648

R;Button, L.L.; McMaster, W.R.

J. Exp. Med. 171, 589, 1990

A;Reference number: PL0221

A;Contents: erratum

A;Accession: PL0221

A;Molecule type: DNA

A;Residues: 1-802 <BU>

A;Cross-references: GB:Y00647; NID:99554; PIDN:CAA86673.1; PID:g9555

A;Note: this is a revision to the sequence from reference A27598

R;Button, L.L.; McMaster, W.R.

J. Exp. Med. 167, 724-729, 1988

A;Title: Molecular cloning of the major surface antigen of *Leishmania*.

A;Reference number: A27598; MUID:88154764; PMID:3346625

A;Accession: A27598

A;Status: significant sequence differences

A;Molecule type: DNA

R;Bouvier, J.; Bordier, C.; Vogel, H.; Reichelt, R.; Etges, R.

Mol. Biochem. Parasitol. 37, 235-246, 1989

A;Title: Characterization of the promastigote surface protease of *Leishmania* as a membr

A;Reference number: A60648; MUID:90114330; PMID:2608099

A;Accession: A60648

A;Molecule type: protein

A;Residues: 101,'E',103-118,'SV',121-123 <BOU>

A;Experimental source: strain LEMS13

R;Schlagenhauf, E.; Etges, R.; Metcalf, P.

submitted to the Brookhaven Protein Data Bank, March 1997

A;Reference number: A68135; PDB:1LMI

A;Contents: annotation; X-ray crystallography, 1.86 angstroms, residues 100-407,412-498.5

A;Note: strain LRC-L119

C;Complex: homodimer

C;Function:

A;Description: catalyzes the hydrolysis of peptide bonds between two hydrophobic residues

A;Note: the activated form can activate the proenzyme form

C;Superfamily: leishmanolysin

C;Keywords: blocked carboxyl end; cell adhesion; glycoprotein; homodimer; hydrolase; lip  
F;1-39/Domain: signal sequence #status predicted <SIG>  
F;40-100/Domain: activation peptide #status predicted <ATP>  
F;101-577/Product: leishmanolysin #status experimental <MAT>  
F;578-602/Domain: carboxyl-terminal propeptide #status predicted <CTP>  
F;48,264,268,334/Binding site: zinc, catalytic (Cys, His, His, His) (inhibited) #status  
F;100-101/Cleavage site: Val-Val (autolytic) #status experimental  
F;125-142,191-230,314-386,393-455,406-425,415-489,466-510,515-565,535-558/Disulfide bond  
F;264,268,334/Binding site: zinc, catalytic (His) (active) #status experimental  
F;265/Active site: Glu #status predicted  
F;300,407/Binding site: carboxylate (Asn) (covalent) #status experimental  
F;577/Modified site: GPI-anchor ethanolamine amidated carboxyl end (Asn) (in mature form

Query Match 11.8%; Score 91; DB 1; Length 602;  
Best Local Similarity 37.5%; Pred. No. 4;  
Matches 27; Conservative 6; Mismatches 15; Indels 24; Gaps 3;  
QY 63 ARKYDQLVTRVVTHEMAHA-----GQADNSTLTQTGFRNNATIDQMNKNSDITV--- 115  
Db 251 ASRYDQLVTRVVTHEMAHALGFGSGPFEDARIV-----ANVPNVRGKNFDVPVINS 301  
QY 116 -----GQYG 119  
Db 302 STAVAKAREQYG 313

Search completed: August 2, 2004, 14:56:23  
Job time : 9.4 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: August 2, 2004, 14:36:12 ; Search time 5.3 Seconds  
(without alignments)  
1493.508 Million cell updates/sec

Title: US-09-543-407-20

Perfect score: 774

Sequence: 1 MKLLKVAFAAIWGSALA.....DSSVMVQVGFGNATNQY 151

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 141681 seqs, 52070155 residues

Total number of hits satisfying chosen parameters: 141681

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt\_42.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	689	89.0	151	1 CSGA SALTY	P55225 salmonella
2	528	68.2	151	1 CSGA ECOLI	P28307 escherichia
3	506.5	65.4	152	1 CSGA ECOLI	Q93u24 escherichia
4	96.5	12.5	151	1 CSGA ECOLI	P38828 escherichia
5	92.5	12.0	1567	1 ICEN_XANCT	P18127 xanthomonas
6	91.5	11.8	1322	1 ICEA PANAN	P20469 pantoea ana
7	91	11.8	602	1 GP63 LEIMA	P08148 leishmania
8	88.5	11.4	151	1 CSGB SALTY	Q8z7m3 salmonella
9	88.5	11.4	151	1 CSGB SALTY	P55226 salmonella
10	88	11.4	590	1 GP63 LEIDO	P23223 leishmania
11	88	11.4	599	1 GP63 LEICH	P15706 leishmania
12	88	11.4	646	1 GP63 LEIME	P43150 leishmania
13	86.5	11.2	369	1 PST3 MYCAV	Q9KK89 mycobacteri
14	84.5	10.9	188	1 FMF2 ECOLI	P02972 escherichia
15	84.5	10.9	1034	1 ICEN PANAN	Q47879 pantoea ana
16	84.5	10.9	1258	1 ICEN ESWHE	P16239 erwinia her
17	83	10.7	678	1 YF48 MYCTU	Q10778 mycobacteri
18	82.5	10.7	592	1 CEA_CITFR	P04480 citrobacter
19	82.5	10.7	641	1 IMD_ARTGO	Q4052 artobacter
20	82.5	10.7	1028	1 OVO DROME	P15121 drosophila
21	81.5	10.5	392	1 HME1 HUMAN	Q05925 homo sapien
22	81.5	10.5	1317	1 N145 YEAST	P49687 saccharomyc
23	81	10.5	1656	1 OMPE RICJA	O06653 r outer mem
24	80.5	10.4	1210	1 ICEN PSEFL	P09815 pseudomonas
25	78.5	10.1	472	1 SACE BACAM	P21130 bacillus am
26	78.5	10.1	487	1 Y442 MYCTU	P27342 caulobacter
27	78	10.1	147	1 HFAA CAUCR	P28314 coprinus ci
28	78	10.1	363	1 PER COPCI	P04273 mesocricetu
29	77.5	10.0	254	1 PRIO MESAU	P28313 arthronyces
30	77.5	10.0	364	1 PER ARTRA	P29579 methanobact
31	77.5	10.0	499	1 YP25 METTF	O86343 mycobacteri
32	77	9.9	370	1 PST3 MYCTU	Q95211 cryotolagus
33	76	9.8	252	1 PRIO RABIT	

34	76	9.8	504	1 FLIC SALMC	Q06981 salmonella
35	76	9.8	955	1 FRU DROME	Q8in81 drosophila
36	76	9.8	1115	1 TEC2 CHIRE	Q8vx33 chlamydomon
37	75.5	9.8	254	1 PRIO CRIGR	Q80506 cricetus
38	75.5	9.8	576	1 CEA7 ECOLI	Q47112 escherichia
39	75.5	9.8	1196	1 ICEV_PSESX	O33479 pseudomonas
40	75	9.7	172	1 CH18 DROME	P07184 drosophila
41	75	9.7	389	1 ACUC STAAM	Q99tc9 staphylococ
42	74.5	9.6	163	1 HCV NATPH	P39442 natrononcha
43	74.5	9.6	370	1 PGLI PENOL	Q39834 penicillium
44	74.5	9.6	663	1 DUS8 MOUSE	Q09112 mus musculu
45	74.5	9.6	760	1 YBIL ECOLI	P75780 escherichia

#### ALIGNMENTS

RESULT 1					
CSGA SALTY					
ID	CSGA SALTY	STANDARD;	PRT;	151 AA.	
AC	P55225;				
DT	01-OCT-1996 (Rel. 34, Created)				
DT	01-OCT-1996 (Rel. 34, Last sequence update)				
DT	10-OCT-2003 (Rel. 42, Last annotation update)				
DE	Major curlin subunit precursor (Fimbrin SEF17).				
GN	CSGA OR AGFA OR STM1144 OR STY1181 OR TI1776.				
OS	Salmonella typhimurium,				
OS	Salmonella typhi, and				
OS	Salmonella enteritidis.				
OC	Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;				
OC	Enterobacteriaceae; Salmonella.				
OX	NCBI_TaxID=602, 601, 592;				
RN	[1]				
RP	SEQUENCE FROM N.A.				
RC	SPECIES=S.typhimurium; STRAIN=SR-11;				
RX	MEDLINE=98117058; PubMed=9457880;				
RA	Romling U., Bian Z., Hammar M., Sierralta W.D., Normark S.;				
RT	"Curli fibers are highly conserved between Salmonella typhimurium and				
RT	Escherichia coli with respect to operon structure and regulation.;"				
RL	J. Bacteriol. 180:722-731(1998).				
RN	[2]				
RP	SEQUENCE FROM N.A.				
RC	SPECIES=S.typhimurium; STRAIN=LT2 / SGSC1412 / ATCC 700720;				
RX	MEDLINE=21534948; PubMed=11677609;				
RA	McClelland M., Sanderson K.E.; Spieth J., Clifton S.W., Latreille P.,				
RA	Courtney L., Porwollik S., Ali J., Dante M., Du F., Hou S., Layman D.,				
RA	Leonard S., Nguyen C., Scott K., Holmes A., Grewal N., Mulvaney E.,				
RA	Ryan E., Sun H., Flores L., Miller W., Stoneking T., Nhan M.,				
RA	Waterston R., Wilson R.K.;				
RT	"Complete genome sequence of Salmonella enterica serovar Typhimurium				
RT	LT2.;"				
RL	Nature 413:852-856(2001).				
RN	[3]				
RP	SEQUENCE FROM N.A.				
RC	SPECIES=S.typhi; STRAIN=CT18;				
RX	MEDLINE=21534947; PubMed=11677608;				
RA	Parthill J., Dougan G., James K.D., Thomson N.R., Pickard D., Wain J.,				
RA	Churcher C., Mungall K.L., Bentley S.D., Holden M.T.G., Sebaihia M.,				
RA	Baker S., Basham D., Brooks K., Chillingworth T., Cornetson P.,				
RA	Cronin A., Davis P., Davies R.M., Dowd L., White N., Farrar J.,				
RA	Feltwell T., Hamlin N., Haque A., Hien T.T., Holtroyd S., Jagels K.,				
RA	Krogh A., Larsen T.S., Leather S., Moule S., O'Gaora P., Parry C.,				
RA	Quail M.A., Rutherford K., Simmonds M., Skelton J., Stevens K.,				
RA	Whitehead S., Barrall B.G.;				
RT	"Complete genome sequence of a multiple drug resistant Salmonella				
RT	enterica serovar Typhi CT18.;"				
RL	Nature 413:848-852(2001).				
RN	[4]				
RP	SEQUENCE FROM N.A.				
RC	SPECIES=S.typhi; STRAIN=TY2 / ATCC 700931;				
RX	MEDLINE=22531367; PubMed=12644504;				
RA	Deng W., Liou S.-R., Plunkett G. III, Mayhew G.F., Rose D.J.,				
RA	Burland V., Kodoyianni V., Schwartz D.C., Blattner F.R.;				

"Comparative genomics of Salmonella enterica serovar Typhi strains Ty2 and CT18.";

J. Bacteriol. 185:2330-2337 (2003).

[5]

SEQUENCE FROM N.A.

SPECIES-S: enteritidis; STRAIN=27655-3B;

MEDLINE=96146512; PubMed=8550497;

Collinson S.K., Clouthier S.C., Doran J.L., Baner P.A., Kay W.W.;

Salmonella enteritidis agfBAC operon encoding thin, aggregative

fimbriae.";

J. Bacteriol. 178:662-667 (1996).

[6]

SEQUENCE OF 21-151 FROM N.A.

SPECIES-S: enteritidis; STRAIN=27655-3B;

MEDLINE=94013373; PubMed=8104955;

Doran J.L., Collinson S.K., Burian J., Sarlos G., Todd E.C.D.,

Munro C.K., Kay C.M., Baner P.A., Peterkin P.I., Kay W.W.;

"PNA-based diagnostic tests for Salmonella species targeting agfA,

the structural gene for thin, aggregative fimbriae.";

J. Clin. Microbiol. 31:2263-2273 (1993).

[7]

SEQUENCE OF 21-33.

SPECIES-S: enteritidis; STRAIN=27655-3B;

MEDLINE=91310586; PubMed=1677357;

Collinson S.K., Emeody L., Mueller K.-M., Trust T.J., Kay W.W.;

"Purification and characterization of thin, aggregative fimbriae from

Salmonella enteritidis.";

J. Bacteriol. 173:4773-4781 (1991).

CC -!- FUNCTION: CURLIN IS THE STRUCTURAL SUBUNIT OF THE CURLI. CURLI ARE

COILED SURFACE STRUCTURES THAT ASSEMBLE PREFERENTIALLY AT GROWTH

TEMPERATURES BELOW 37 DEGREES CELSIUS. CURLI CAN BIND TO

FIBRONECTIN

CC -!- SIMILARITY: BELONGS TO THE CSGA/CSGB FAMILY.

CC

-----

THIS SWISS-PROT entry is copyright. It is produced through a collaboration

between the Swiss Institute of Bioinformatics and the EMBL outstation

the European Bioinformatics Institute. There are no restrictions on its

use by non-profit institutions as long as its content is in no way

modified and this statement is not removed. Usage by and for commercial

entities requires a license agreement (See <http://www.isb-sib.ch/announce/>

or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

-----

EMBL; AJ002301; CAA05317.1; -

EMBL; AE008749; AAL20074.1; -

EMBL; AL627269; CAD08268.1; -

EMBL; AE016840; AAC069399.1; -

EMBL; U43280; AAC43599.1; -

PIR; JC6039; JG6039.

DR StyGene; SG10608; csqa.

KW Fimbria; Signal; Complete proteome.

FT SIGNAL 1 20

FT CHAIN 21 151 MAJOR CURLIN SUBUNIT.

FT CONFLICT 134 151 SVMVRQVGFNNATANYQ -> DSYTQVAS (IN

REF. 6).

FT SEQUENCE 151 AA; 1305 MW; 87DAC0D16B621359 CRC64;

Query Match 89.0%; Score 689; DB 1; Length 151;

Best Local Similarity 90.7%; Pred. No. 3.9e-53;

Matches 137; Conservative 3; Mismatches 11; Indels 0; Gaps 0;

QY 1 MKLKVAAFAIVVSGSALAGVWPQGGGNGHNGGNSGPDSTLSIYQVGSANAALALQ 60

DB 1 MKLKVAAFAIVVSGSALAGVWPQGGGNGHNGGNSGPDSTLSIYQVGSANAALALQ 60

QY 61 SDARKYQDLVTRVVTHEMAHAGQADNSTIELTQNGFRNNATIDQNAKNSDITVQYGG 120

DB 61 SDARKSETTITQSGYNGGADVQGGADNSTIELTQNGFRNNATIDQNAKNSDITVQYGG 120

QY 121 NNAALVNOTASDSVVRQVGFNNATANYQ 151

DB 121 NNAALVNOTASDSVVRQVGFNNATANYQ 151

RESULT 2

CSGA\_ECOLI STANDARD; PRT; 151 AA.

ID \_CSGA\_ECOLI

AC P28307;

DT 01-DEC-1992 (Rel. 24, Created)

DT 01-OCT-1996 (Rel. 34, Last sequence update)

DT 28-FEB-2003 (Rel. 41, Last annotation update)

DE Major curlin subunit precursor.

GN CSGA OR B1042.

OS Escherichia coli.

OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;

OC Enterobacteriaceae; Escherichia.

OX NCBI\_TaxID=562;

XX [1]

RP SEQUENCE FROM N.A.

RC STRAIN=K12 / W3110;

RX MEDLINE=93211294; PubMed=8459772;

RA Olsen A., Arngvist A.;

RT "The RpoS sigma factor relieves H-NS-mediated transcriptional

repression of csqa, the subunit gene of fibronectin-binding curli in

Escherichia coli.";

RL Mol. Microbiol. 7:523-536 (1993).

[2]

RP SEQUENCE FROM N.A.

RC STRAIN=K12 / MC4100;

RX MEDLINE=96414469; PubMed=8817489;

RA Hammar M., Arngvist A., Bian Z., Olsen A., Normark S.;

RT "Expression of two csg operons is required for production of

fibronectin- and congo red-binding curli polymers in Escherichia coli

K-12.";

RL Mol. Microbiol. 18:661-670 (1995).

[3]

RP SEQUENCE FROM N.A.

RC STRAIN=K12 / MG1655;

RX MEDLINE=97426617; PubMed=9278503;

RA Blatter F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,

Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,

Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,

Mau B., Shao Y.;

RT "The complete genome sequence of Escherichia coli K-12.";

Science 277:1453-1474 (1997).

[4]

RP SEQUENCE FROM N.A.

RC STRAIN=K12;

RX MEDLINE=97061202; PubMed=8905232;

Oshima T., Aiba H., Baba T., Fujita K., Hayashi K., Honjo A.,

Ikenoto K., Inada T., Itoh T., Kajihara M., Kanai K., Kashimoto K.,

Kimura S., Kitagawa M., Makino K., Masuda S., Miki T., Mizobuchi K.,

Mori H., Motomura K., Nakamura Y., Nishimoto H., Nishio Y., Saito N.,

Sampeki G., Seki Y., Tagami H., Takemoto K., Wada C., Yamamoto Y.,

Yano M., Horiuchi T.;

RT "A 718-Kb DNA sequence of the Escherichia coli K-12 genome

corresponding to the 12.7-28.0 min region on the linkage map.";

DNA Res. 3:137-155 (1996).

[5]

RP SEQUENCE OF 21-40.

RC STRAIN=K12 / YMEI;

RX MEDLINE=93023873; PubMed=1357528;

RA Arngvist A., Olsen A., Pfeiffer J., Russell D.G., Normark S.;

RT "The Crl protein activates cryptic genes for curli formation and

fibronectin binding in Escherichia coli HB101.";

RL Mol. Microbiol. 6:2443-2452 (1992).

[6]

RP SEQUENCE OF 21-31.

RX MEDLINE=91310586; PubMed=1677357;

Collinson S.K., Emeody L., Trust T.J., Kay W.W.;

RT "Purification and characterization of thin, aggregative fimbriae from

Salmonella enteritidis.";

RL J. Bacteriol. 173:4773-4781 (1991).

CC -!- FUNCTION: CURLIN IS THE STRUCTURAL SUBUNIT OF THE CURLI. CURLI ARE

COILED SURFACE STRUCTURES THAT ASSEMBLE PREFERENTIALLY AT GROWTH

TEMPERATURES BELOW 37 DEGREES CELSIUS. CURLI CAN BIND TO

FIBRONECTIN.

CC



```

[3]
SEQUENCE FROM N.A.
RN RP STRAIN=O157:H7 / RIMD 0509952;
RC MEDLINE=21156231; PubMed=11258796;
RX Hayaishi T., Makino K., Ohnishi M., Kurokawa K., Ishii K., Yokoyama K.,
RA Han C.-G., Ohtsubo E., Nakayama K., Murata T., Tanaka M., Tobe T.,
RA Lida T., Takami H., Honda T., Sasaki K., Ogasawara N., Yasunaga T.,
RA Kuhara S., Shiba T., Hattori M., Shinagawa H.;
RT "Complete genome sequence of enterohemorrhagic Escherichia coli
RL O157:H7 and genomic comparison with a laboratory strain K-12."
RL DNA Res. 8:11-22(2001).
CC -!- FUNCTION: CURLIN IS THE STRUCTURAL SUBUNIT OF THE CURLIN. CURLI ARE
CC COILED SURFACE STRUCTURES THAT ASSEMBLE PREFERENTIALLY AT GROWTH
CC TEMPERATURES BELOW 37 DEGREES CELSIUS. CURLI CAN BIND TO
CC FIBRONECTIN.
CC -!- SIMILARITY: BELONGS TO THE CSGA/CSGB FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; AF275733; AAK53212.1; -
CC EMBL; A3005315; AAG55788.1; -
CC EMBL; AP002554; BAB34843.1; -
CC FIR; D90806; D90806.
CC FIR; H85665; H85665.
CC KW Fimbria; Signal; Complete proteome.
CC SIGNAL 1 20
CC CHAIN 21 152 MAJOR CURLIN SUBUNIT.
CC SEQUENCE 152 AA; 15099 MW; EE2D2D94DDE91243 CRC64;

Query Match 65.4%; Score 506.5; DB 1; Length 152;
Best Local Similarity 67.1%; Pred. No. 2.6e-37;
Matches 102; Conservative 21; Mismatches 28; Indels 1; Gaps 1;

QY 1 MKLLKVAFAAIVVSGSALAGVVPQW-CGGGNHNGGNSGDPSTLSIYQYGSANAALAL 59
DB 1 MKLLKVAATAAVFSSALAGVVPYQYGGGNGHGGGNSGPNSEINIVYQYGGNSALAL 60
QY 60 QSDARKYDYLQVTRVTHVAHAGQAGDNSTIELTQNGPRNATIDOWNAKNSDITVGYQG 119
DB 61 QADARNSDLTITTHGGCGNGADYVQGSDDSSIDLTQRFQNGSATLDQWNGKDSHMTVKQFG 120
QY 120 GNNAAVNOTASDSSVMVTVQVFGNNATANQY 151
DB 121 GNGAAVNOTASNSTVNVTVQVFGNNATAHQY 152

RESULT 4
CSGB ECOLI
ID _CSGB_ECOLI STANDARD; PRT; 151 AA.
AC P39828;
DT 01-FEB-1995. (Rel. 31, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Minor curlin subunit precursor.
GN CSGB OR B1041 OR Z1675 OR ECS1419.
OS Escherichia coli, and
OS Escherichia coli O157:H7.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Escherichia.
OC NCBI_TaxID=562, 83334;
[1]
SEQUENCE FROM N.A.
RN RP STRAIN=K12 / MC4100;
RC MEDLINE=96414468; PubMed=8817489;
RX Hammar M., Arngvist A., Bian Z., Olsen A., Normark S.;
RA "Expression of two csg operons is required for production of
RT fibronectin- and Congo red-binding curli polymers in Escherichia coli
```

DR

SQ SEQUENCE 1567 AA; 152548 MW; C8B451D959ECAD63 CRC64;  
Query Match 12.0%; Score 92.5; DB 1; Length 1567;  
Best Local Similarity 25.8%; Pred. No. 3.1;  
Matches 46; Conservative 27; Mismatches 56; Indels 49; Gaps 11;  
QY 14 VSGSALAG-----VVPQWGG-----GGNHN-----GGGNSGPDSTLSIVQVGSANAALAL 59  
DB 205 VVGSILTGADQSRVLVAGYSTAGDHSLLIAGYSTGTAGSDSSI-LAGYSTQTAAGR 263  
QY 60 QSDARKYDQLV-----RVVTHEMAHAGQGANSTIELTQNGFRNNAT----- 103  
DB 264 STLTAGYGTQTAAQSGSRLTSGYGTATSGDSAVI-----SGYGTQTAGSESSLTGYG 319  
QY 104 -DOWNAKNSDITVGVG-----GNNALV-----NQTASDSSVMVVRQVGFNNATNQ 150  
DB 320 STQTARKGSDITAG-YGSTGTAGSDSALIAGYSTGTAGSESSLT--AGYSTQTARK 374  
RESULT 6  
ID ICEA PANAN STANDARD; PRT; 1322 AA.  
AC P20469; 1991 (Rel. 17, Created)  
DT 01-FEB-1991 (Rel. 17, Last sequence update)  
DT 16-OCT-2001 (Rel. 40, Last annotation update)  
DE Ice nucleation protein InaA.  
GN INAA  
OS Pantoea ananas (Erwinia uredovora).  
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;  
OC Enterobacteriaceae; Pantoea.  
OX NCBI\_TaxID=553;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=90092494; PubMed=2559095;  
RA Abe K., Watabe S., Emori Y., Watanabe M., Arai S.;  
RT "An ice nucleation active gene of Erwinia ananas. Sequence similarity  
RT to those of Pseudomonas species and regions required for ice  
RT nucleation activity.";  
RL FEBS Lett. 258:297-300(1989).  
CC -!- FUNCTION: Ice nucleation proteins enable bacteria to nucleate  
CC crystallization in supercooled water.  
CC -!- SUBCELLULAR LOCATION: Outer membrane (By similarity).  
CC -!- DOMAIN: CONTAINS MANY IMPERFECT REPEATS OF THE CONSENSUS  
CC OCTAPEPTIDE A-G-Y-G-S-T-X-T; FURTHER ON A 16-RESIDUE AND A  
CC REGIONAL 48-RESIDUE PERIODICITY IS SUPERIMPOSED.  
CC -!- SIMILARITY: Belongs to the bacterial ice nucleation protein  
CC family.  
CC  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC  
CC EMBL; X17316; CAA35194.1; -  
DR PIR; S07053; S07053.  
DR HSP; P06620; IINA.  
DR InterPro; IPR000258; Ice nucleatn.  
DR Pfam; PF00818; Ice nucleation; 69.  
DR PRINTS; PR00327; ICNUCLEATN.  
DR PROSITE; PS00314; ICE\_NUCLEATION; 49.  
KW Ice nucleation; Repeat; Outer membrane.  
FT DOMAIN 162 1281 OCTAPEPTIDE PERIODICITY.  
SQ SEQUENCE 1322 AA; 131094 MW; 89B0EE24AA637039 CRC64;  
Query Match 11.8%; Score 91.5; DB 1; Length 1322;  
Best Local Similarity 29.1%; Pred. No. 3.1;  
Matches 37; Conservative 19; Mismatches 22; Indels 49; Gaps 9;  
QY 34 GGNSSGPDSTLSIVQVGSANAALALQSDARKYDQLVTRVTHEMAHAGQGANSTIELT 93

DB 933 GSTSTAGPDSLSL-LAGYSTQTA-----GYSILT-----AGYGS-----T 967  
QY 94 QNGFRNNATIDQNNAKNSDITVGVG-----GNNALV-----NQTASDSSVMVVRQVGF 143  
DB 968 QTG-----QENSDLTTG-YGSTAGYESSLIAGYSTQTASFKSLM--AGYG 1013  
QY 144 NNATANQ 150  
DB 1014 SSQTARE 1020  
RESULT 7  
ID GP63 LEIMA STANDARD; PRT; 602 AA.  
AC P08148; P15906;  
DT 01-AUG-1988 (Rel. 08, Created)  
DT 01-APR-1990 (Rel. 14, Last sequence update)  
DT 10-OCT-2003 (Rel. 42, Last annotation update)  
DE Leishmanolysin precursor (EC 3.4.24.36) (Cell surface protease)  
DE (Major surface glycoprotein) (GP63 protein) (Promastigote surface  
DE endopeptidase).  
GN GP63.  
OS Leishmania major.  
OC Eukaryota; Euglenozoa; Kinetoplastida; Trypanosomatidae; Leishmania.  
OX NCBI\_TaxID=5664;  
RN [1]  
RP SEQUENCE FROM N.A., AND SEQUENCE OF 101-123.  
RX MEDLINE=98154764; PubMed=3346625;  
RA Button L.L., McWaster W.R.;  
RT "Molecular cloning of the major surface antigen of leishmania.";  
RL J. Exp. Med. 167:724-729(1988).  
RN [2]  
RP REVISIONS.  
RA Button L.L., McMaster W.R.;  
RL J. Exp. Med. 171:589-589(1990).  
RN [3]  
RP GPI-ANCHOR.  
RX MEDLINE=91009116; PubMed=2145267;  
RA Schneider P., Ferguson M.A.J., McConville M.J., Mehler A.,  
RA Homans S.W., Bordier C.;  
RT "Structure of the glycosyl-phosphatidylinositol membrane anchor of  
RT the Leishmania major promastigote surface protease.";  
RL J. Biol. Chem. 265:16955-16964(1990).  
RN [4]  
RP X-RAY CRYSTALLOGRAPHY (2.6 ANGSTROMS).  
RX MEDLINE=95406217; PubMed=7675788;  
RA Schlagenhauf E., Etges R., Metcalf P.;  
RT "Crystallization and preliminary X-ray diffraction studies of  
RT leishmanolysin, the major surface metalloproteinase from Leishmania  
RT major.";  
RN Proteins 22:58-66(1995).  
RN [5]  
RP X-RAY CRYSTALLOGRAPHY (1.86 ANGSTROMS).  
RX MEDLINE=98416698; PubMed=9739094;  
RA Schlagenhauf E., Etges R., Metcalf P.;  
RT "The crystal structure of the Leishmania major surface proteinase  
RT leishmanolysin.";  
RL Structure 6:1035-1046(1998).  
CC -!- FUNCTION: Has an integral role during the infection of macrophages  
CC in the mammalian host.  
CC -!- CATALYTIC ACTIVITY: Preference for hydrophobic residues at P1 and  
CC P1' and basic residues at P2 and P3'. A model nonapeptide is  
CC Cleaved at Ala-Tyr-Leu-Lys-Lys-  
CC -!- COFACTOR: Binds 1 zinc ion per subunit.  
CC -!- SUBCELLULAR LOCATION: Attached to the membrane by a GPI-anchor.  
CC -!- PTM: THE PHOSPHATIDYLINOSITOL MOIETY OF THE GPI-ANCHOR CONTAINS A  
CC FULLY SATURATED, UNBRANCHED 1-O-ALKYL CHAIN (MAINLY C24:0) AND A  
CC MIXTURE OF FULLY SATURATED UNBRANCHED 2-O-ACYL CHAINS (C12:0,  
CC C14:0, C16:0, AND C18:0).  
CC -!- SIMILARITY: Belongs to peptidase family M8.  
CC  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----

EMBL; Y00647; CAA68673.1; -  
PIR; P0221; P0221.  
PDB; 1LML; 17-SEP-97.  
DR MEROPS; M08.001; -  
DR InterPro; IPR006025; Pept M Zn BS.  
DR InterPro; IPR001577; Peptidase\_M8.  
DR Pfam; PF01457; Peptidase\_M8\_1.  
DR PRINTS; PR00782; LSHMANOLYSIN.  
DR PROSITE; PS00142; ZINC\_PROTEASE; 1.  
KW Hydrolase; Metalloprotease; Glycoprotein; Metal-binding; Zinc;  
KW Zymogen; Signal; Cell adhesion; 3D-structure; Lipoprotein.  
FT SIGNAL 1 39 POTENTIAL.  
FT PROPEP 40 100 ACTIVATION PEPTIDE.  
FT CHAIN 101 577 LEISHMANOLYSIN.  
FT PROPEP 578 602 REMOVED IN MATURE FORM.  
FT METAL 264 264 ZINC (CATALYTIC).  
FT ACT\_SITE 265 265 ZINC (CATALYTIC).  
FT METAL 268 268 ZINC (CATALYTIC).  
FT METAL 334 334 ZINC (CATALYTIC).  
FT DISULFID 125 142 ZINC (CATALYTIC).  
FT DISULFID 191 230 ZINC (CATALYTIC).  
FT DISULFID 314 386  
FT DISULFID 393 455  
FT DISULFID 406 425  
FT DISULFID 415 489  
FT DISULFID 466 510  
FT DISULFID 515 565  
FT DISULFID 535 588  
FT CARBOHYD 300 300 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 407 407 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT LIPID 577 577 GPI-anchor amidated asparagine.  
FT STRAND 101 102  
FT STRAND 107 108  
FT STRAND 111 114  
FT HELIX 116 119  
FT TURN 121 122  
FT TURN 128 129  
FT STRAND 131 133  
FT STRAND 139 141  
FT HELIX 144 146  
FT HELIX 150 158  
FT TURN 159 159  
FT TURN 160 169  
FT TURN 170 171  
FT STRAND 172 174  
FT STRAND 177 178  
FT STRAND 180 181  
FT TURN 189 190  
FT HELIX 191 193  
FT HELIX 198 202  
FT TURN 203 203  
FT STRAND 205 206  
FT STRAND 210 215  
FT STRAND 221 222  
FT STRAND 226 232  
FT TURN 234 235  
FT STRAND 238 244  
FT HELIX 247 249  
FT HELIX 256 269  
FT TURN 270 271  
FT HELIX 274 279  
FT TURN 280 281  
FT STRAND 283 286  
FT HELIX 289 291  
FT STRAND 296 296  
FT HELIX 302 312

FT TURN 313 313  
FT TURN 315 316  
FT STRAND 320 322  
FT TURN 328 332  
FT STRAND 335 335  
FT TURN 337 339  
FT TURN 341 342  
FT STRAND 343 343  
FT TURN 344 345  
FT STRAND 353 353  
FT HELIX 356 364  
FT TURN 365 366  
FT STRAND 369 370  
FT HELIX 372 374  
FT TURN 380 383  
FT HELIX 386 390  
FT STRAND 394 395  
FT TURN 396 397  
FT STRAND 398 399  
FT TURN 402 404  
FT STRAND 413 414  
FT TURN 417 418  
FT STRAND 421 425  
FT STRAND 428 429  
FT HELIX 435 437  
FT TURN 443 444  
FT STRAND 445 446  
FT TURN 450 454  
FT STRAND 458 465  
FT TURN 466 467  
FT HELIX 470 472  
FT TURN 475 477  
FT HELIX 478 480  
FT TURN 485 486  
FT STRAND 487 494  
FT STRAND 496 496  
FT STRAND 506 516  
FT TURN 517 520  
FT STRAND 521 525  
FT TURN 527 528  
FT STRAND 533 534  
FT TURN 537 538  
FT STRAND 540 542  
FT HELIX 543 545  
FT TURN 546 546  
FT STRAND 550 550  
FT TURN 552 553  
FT STRAND 555 557  
FT HELIX 561 565  
FT TURN 566 567  
FT HELIX 569 572  
FT TURN 573 573  
SQ SEQUENCE 602 AA; 63953 MW; 982EF3245D87C43E CRC64;

Query Match 11.8%; Score 91; DB 1; Length 602;  
Best Local Similarity 37.5%; Pred. No. 1.4;  
Matches 27; Conservative 6; Mismatches 15; Indels 24; Gaps 3;  
QY 63 ARKYDQVLTTRVVTHEMAHA-----GQADNSTIELTQNGFRNATIDQWNAKNSDITV--- 115  
DB 251 ASRYDQVLTTRVVTHEMAHAALGFGSGPPFEDARIV-----ANVPNRGKNTDVPVINS 301  
QY 116 -----GQYG 119  
DB 302 STAVAKAREQYG 313

RESULT 8  
CSGB\_SALTI ID CSGB\_SALTI STANDARD; PRT; 151 AA.  
AC Q827M3;  
DT 28-FEB-2003 (Rel. 41, Created)  
DT 28-FEB-2003 (Rel. 41, Last sequence update)

DT 10-OCT-2003 (Rel. 42, Last annotation update)  
 DE Minor curlin subunit precursor (Fimbrin SEF17 minor subunit).  
 GN CSGB OR STYL180 OR TL777.  
 OS Salmonella typhimurium, and  
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;  
 OC Enterobacteriaceae; Salmonella.  
 OX NCBI\_TaxID=601;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=CT18;  
 RX MEDLINE=21534947; PubMed=11677608;  
 RA Parikh J., Dougan G., James K.D., Thomson N.R., Pickard D., Wain J.,  
 RA Churcher C., Mungall K.L., Bentley S.D., Holden M.T.G., Sebahia M.,  
 RA Baker S., Basham D., Brooks K., Chillingworth T., Conerton P.,  
 RA Cronin A., Davis P., Davies R.M., Dowd L., White N., Farrar J.,  
 RA Fellwell T., Hamlin N., Haque A., Hien T.T., Holroyd S., Jagels K.,  
 RA Krogh A., Larsen T.S., Leather S., Moulé S., O'Gaora P., Parry C.,  
 RA Quail M.A., Rutherford K., Simmonds M., Skelton J., Stevens K.,  
 RA Whitehead S., Barrrell B.G.;  
 RT "Complete genome sequence of a multiple drug resistant Salmonella  
 RT enterica serovar Typhi CT18.";  
 RL Nature 413:848-852(2001).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=Ty2 / ATCC 700931;  
 RX MEDLINE=22531367; PubMed=12644504;  
 RA Deng W., Liou S.-R., Plunkett G. III, Mayhew G.P., Rose D.J.,  
 RA Burland V., Kodoyianni V., Schwarz D.C., Blattner P.R.;  
 RT "Comparative genomics of Salmonella enterica serovar Typhi strains Ty2  
 RT and CT18.";  
 RL J. Bacteriol. 185:2330-2337(2003).  
 CC -!- FUNCTION: CURLIN IS THE STRUCTURAL SUBUNIT OF THE CURLI. CURLI ARE  
 CC COILED SURFACE STRUCTURES THAT ASSEMBLE PREFERENTIALLY AT GROWTH  
 CC TEMPERATURES BELOW 37 DEGREES CELSIUS. CURLI CAN BIND TO  
 CC FIBRONECTIN. THE MINOR SUBUNIT IS THE NUCLEATION COMPONENT OF  
 CC CURLIN MONOMERS.  
 CC -!- SIMILARITY: BELONGS TO THE CSGA/CSGB FAMILY.  
 CC  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC  
 DR EMBL; AL627269; CAD08267.1; -;  
 DR EMBL; AF016840; AB069400.1; -;  
 KW Fimbria; Signal; Complete proteome.  
 FT SIGNAL 21 POTENTIAL.  
 FT CHAIN 22 151 MINOR CURLIN SUBUNIT.  
 SQ SEQUENCE 151 AA; 16254 MW; 161C54326E573495 CRC64;  
 Query Match 11.4%; Score 88.5; DB 1; Length 151;  
 Best Local Similarity 31.0%; Pred. No. 0.48;  
 Matches 26; Conservative 13; Mismatches 42; Indels 3; Gaps 2;  
 QY 69 LVTRVVTHEMAHAGQADNSTIELTQNGFRNATIDOWNAKNSDIT-VQYGGNNALVN 127  
 Db 48 IIGQVGTDSARVRQEGSKLLSVISQEGNNRAKVDQ--AGNYNFAIETQGNANDASIS 105  
 QY 128 QTASDSSVMVRQVGFNNATANY 151  
 Db 106 QSAYGNSAAIIQKSGNKANITQY 129  
 RESULT 9  
 CSGB SALTY STANDARD; PRT; 151 AA.  
 AC P35226; 1996 (Rel. 34, Created)  
 DT 01-OCT-1996 (Rel. 34, Last sequence update)  
 DT 28-FEB-2003 (Rel. 41, Last annotation update)

DE Minor curlin subunit precursor (Fimbrin SEF17 minor subunit).  
 GN CSGB OR AGFB OR STM1143.  
 OS Salmonella typhimurium, and  
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;  
 OC Enterobacteriaceae; Salmonella.  
 OX NCBI\_TaxID=602, 592;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC SPECIES=S.typhimurium; STRAIN=SR-11;  
 RX MEDLINE=98117058; PubMed=9457880;  
 RA Romling U., Bian Z., Hammar M., Sierralta W.D., Normark S.;  
 RT "Curli fibers are highly conserved between Salmonella typhimurium and  
 RT Escherichia coli with respect to operon structure and regulation.";  
 RL J. Bacteriol. 180:722-731(1998).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC SPECIES=S.typhimurium; STRAIN=LT2 / SGSC1412 / ATCC 700720;  
 RX MEDLINE=21534948; PubMed=11677609;  
 RA McClelland M., Sanderson K.E., Spieth J., Clifton S.W., Latreille P.,  
 RA Courtney L., Porwollik S., Ali J., Dante M., Du F., Hou S., Layman D.,  
 RA Leonard S., Nguyen C., Scott K., Holmes A., Grewal N., Mulvaney E.,  
 RA Ryan E., Sun H., Florea L., Miller W., Stoneking T., Nhan M.,  
 RA Waterston R., Wilson R.X.;  
 RT "Complete genome sequence of Salmonella enterica serovar Typhimurium  
 RT LT2.";  
 RL Nature 413:852-856(2001).  
 RN [3]  
 RP SEQUENCE FROM N.A.  
 RC SPECIES=S. enteritidis; STRAIN=27655-3B;  
 RX MEDLINE=96146512; PubMed=8550497;  
 RA Collinson S.K., Clouthier S.C., Doran J.L., Banser P.A., Kay W.W.;  
 RT "Salmonella enteritidis agfBAC operon encoding thin, aggregative  
 RT fimbriae.";  
 RL J. Bacteriol. 178:662-667(1996).  
 CC -!- FUNCTION: CURLIN IS THE STRUCTURAL SUBUNIT OF THE CURLI. CURLI ARE  
 CC COILED SURFACE STRUCTURES THAT ASSEMBLE PREFERENTIALLY AT GROWTH  
 CC TEMPERATURES BELOW 37 DEGREES CELSIUS. CURLI CAN BIND TO  
 CC FIBRONECTIN. THE MINOR SUBUNIT IS THE NUCLEATION COMPONENT OF  
 CC CURLIN MONOMERS.  
 CC -!- SIMILARITY: BELONGS TO THE CSGA/CSGB FAMILY.  
 CC  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC  
 DR EMBL; AJ002301; CA05316.1; -;  
 DR EMBL; AE008749; AAL20073.1; -;  
 DR EMBL; U43280; AAC43598.1; -;  
 DR PIR; JC6040; JC6040.  
 KW StyGene; SG10609; csGB.  
 KW Fimbria; Signal; Complete proteome.  
 FT SIGNAL 21 POTENTIAL.  
 FT CHAIN 22 151 MINOR CURLIN SUBUNIT.  
 SQ SEQUENCE 151 AA; 16182 MW; C0FC5430B6DD361D CRC64;  
 Query Match 11.4%; Score 88.5; DB 1; Length 151;  
 Best Local Similarity 31.0%; Pred. No. 0.48;  
 Matches 26; Conservative 13; Mismatches 42; Indels 3; Gaps 2;  
 QY 69 LVTRVVTHEMAHAGQADNSTIELTQNGFRNATIDOWNAKNSDIT-VQYGGNNALVN 127  
 Db 48 IIGQVGTDSARVRQEGSKLLSVISQEGNNRAKVDQ--AGNYNFAIETQGNANDASIS 105  
 QY 128 QTASDSSVMVRQVGFNNATANY 151  
 Db 106 QSAYGNSAAIIQKSGNKANITQY 129

```

RESULT 10
GP63_LEIDO STANDARD; PRT; 590 AA.
ID ID GP63_LEIDO STANDARD; PRT; 590 AA.
AC P23223;
DT 01-NOV-1991 (Rel. 20, Created)
DT 01-NOV-1991 (Rel. 20, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Leishmanolysin precursor (EC 3.4.24.36) (Cell surface protease)
DE (Major surface glycoprotein) (GP63 protein) (Promastigote surface
DE endopeptidase).
GN GP63.
OS Leishmania donovani.
OC Eukaryota; Euglenozoa; Kinetoplastida; Trypanosomatidae; Leishmania.
OX NCBI_TaxID=5661;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=LV9;
RX MEDLINE=92107220; PubMed=1762629;
RA Webb J.R., Button L.L., McMaster R.W.;
RT "Heterogeneity of the genes encoding the major surface glycoprotein
RT of Leishmania donovani."
RL Mol. Biochem. Parasitol. 48:173-184(1991).
CC -!- FUNCTION: Has an integral role during the infection of macrophages
CC in the mammalian host.
CC -!- CATALYTIC ACTIVITY: Preference for hydrophobic residues at P1 and
CC p1', and basic residues at P2 and P3'. A model nonapeptide is
CC cleaved at -Ala-Tyr-[Leu-Lys-Lys-].
CC -!- COFACTOR: Binds 1 zinc ion per subunit (By similarity).
CC -!- SUBCELLULAR LOCATION: Attached to the membrane by a GPI-anchor.
CC -!- SIMILARITY: Belongs to peptidase family M8.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; M50048; AAA29244.1; -.
CC HSP; P08148; 1LML.
CC MEROPS; M08.001; -.
CC InterPro; IPR006025; Pept M_Zn_BS.
CC Pfam; PF01457; Peptidase M8.
CC PRINTS; PR00782; LSHMANOLYSIN.
CC PROSITE; PS00142; ZINC PROTEASE; 1.
CC Hydrolase; Metalloprotease; Glycoprotein; Metal-binding; Zinc;
KW Zymogen; Signal; Cell adhesion; GPI-anchor; Lipoprotein.
FT SIGNAL 1 39
FT PROPEP 40 87
FT CHAIN 88 565
FT PROPEP 566 590
FT METAL 251 251
FT METAL 252 252
FT ACT_SITE 252 252
FT METAL 321 321
FT METAL 321 321
FT DISULFID 112 129
FT DISULFID 178 217
FT DISULFID 301 373
FT DISULFID 380 443
FT DISULFID 393 412
FT DISULFID 402 477
FT DISULFID 454 498
FT DISULFID 503 553
FT DISULFID 523 546
FT CARBOHYD 287 287
FT LIPID 565 565
FT SEQUENCE 590 AA; 62950 MW; 0FB315D299659F58 CRC64;
Query Match 11.4%; Score 88; DB 1; Length 590;
Best Local Similarity 89.5%; Pred. No. 2.5;

```

Matches 17; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Qy 63 ARYDQLVTRVTHMAHA 81  
|:|||||  
Db 238 ASRYDQLVTRVTHMAHA 256

RESULT 11  
GP63\_LEICH STANDARD; PRT; 599 AA.  
ID GP63\_LEICH STANDARD; PRT; 599 AA.  
AC P15706;  
DT 01-APR-1990 (Rel. 14, Created)  
DT 01-APR-1990 (Rel. 14, Last sequence update)  
DT 10-OCT-2003 (Rel. 42, Last annotation update)  
DE Leishmanolysin precursor (EC 3.4.24.36) (Cell surface protease)  
DE (Major surface glycoprotein) (GP63 protein) (Promastigote surface  
DE endopeptidase).  
GN GP63.  
OS Leishmania chagasi.  
OC Eukaryota; Euglenozoa; Kinetoplastida; Trypanosomatidae; Leishmania.  
OX NCBI\_TaxID=44271;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=90205976; PubMed=2320059;  
RA Miller R.A., Reed S.G., Parsons M.;  
RT "Leishmania gp63 molecule implicated in cellular adhesion lacks an  
RT Arg-Gly-Asp sequence."  
RL Mol. Biochem. Parasitol. 39:267-274(1990).  
RN [2]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=92112918; PubMed=1370484;  
RA Ramamoorthy R., Donelson J.E., Paetz K.E., Maybodi M., Roberts S.C.,  
RA Wilson M.E.;  
RT "Three distinct RNAs for the surface protease GP63 are differentially  
RT expressed during development of Leishmania donovani chagasi  
RT promastigotes to an infectious form."  
RL J. Biol. Chem. 267:1888-1895(1992).  
CC -!- FUNCTION: Has an integral role during the infection of macrophages  
CC in the mammalian host.  
CC -!- CATALYTIC ACTIVITY: Preference for hydrophobic residues at P1 and  
CC p1', and basic residues at P2 and P3'. A model nonapeptide is  
CC cleaved at -Ala-Tyr-[Leu-Lys-Lys-].  
CC -!- COFACTOR: Binds 1 zinc ion per subunit (By similarity).  
CC -!- SUBCELLULAR LOCATION: Attached to the membrane by a GPI-anchor.  
CC -!- SIMILARITY: Belongs to peptidase family M8.  
CC  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC  
CC EMBL; M80672; AAA29238.1; -.  
CC EMBL; M28527; AAA29235.1; -.  
CC PIR; A43951; A44951.  
CC HSP; P08148; 1LML.  
CC MEROPS; M08.001; -.  
CC InterPro; IPR006025; Pept M\_Zn\_BS.  
CC Pfam; PF01457; Peptidase M8.  
CC PRINTS; PR00782; LSHMANOLYSIN.  
CC PROSITE; PS00142; ZINC PROTEASE; 1.  
CC Hydrolase; Metalloprotease; Glycoprotein; Metal-binding; Zinc;  
KW Zymogen; Signal; Cell adhesion; GPI-anchor; Lipoprotein.  
FT SIGNAL 1 39  
FT PROPEP 40 97  
FT CHAIN 98 574  
FT PROPEP 575 599  
FT METAL 261 261  
FT ACT\_SITE 262 262  
FT METAL 265 265



```

DR EMBL; AF137360; AAF74819.1; -.
DR HSSP; P06128; IA54.
DR InterPro; IPR000437; Prok lipoprot.S.
DR Pfam; PF01547; SBP_bac_1.
DR PROSITE; PS00013; PROKAR LIPOPROTEIN; 1.
KW Phosphate transport; Transport; Membrane; Lipoprotein; Signal;
KW Palmitate.
FT SIGNAL 1 22 POTENTIAL.
FT CHAIN 23 369 PHOSPHATE-BINDING PROTEIN 3.
FT LIPID 23 23 N-palmitoyl cysteine (Potential).
FT LIPID 23 23 S-diacylglycerol cysteine (Potential).
SQ SEQUENCE 369 AA; CB0EA0AC10F463EC CRC64;

Query Match 11.2%; Score 86.5; DB 1; Length 369;
Best Local Similarity 27.8%; Pred. No. 2;
Matches 35; Conservative 21; Mismatches 49; Indels 21; Gaps 6;

QY 1 MKLLKVAFAAIVVSGSALAGVVVPQWGGGNGGNSGPDSTLSIYQYGSANAALALQ 60
DB 1 MKLLNRGAVLSVLSAGLVL-----SGGSDNNGAGAGAGSSSKYSCGK3ALKASG 54
QY 61 SDARKYDQIVTRVW-THMAHAGQADNSTIELTONGFRNATIDQWNAKNSDIYVGQV 119
DB 55 STAQA--NMTREVNAFEQACPCQ-----TLNYTANG--SGAGISEFNGKQTD-----FG 100
QY 120 GUNNAAL 125
DB 101 GDSPL 106

RESULT 14
FME2_ECOLI STANDARD; PRT; 188 AA.
AC P02972;
DT 21-JUL-1986 (Rel. 01, Created)
DT 21-JUL-1986 (Rel. 01, Last sequence update)
DT 15-MAR-2004 (Rel. 43, Last annotation update)
DE F7-2 fimbrial protein precursor (F7-2 pilin).
GN F7-2 OR PAPA OR C3592.
OS Escherichia coli, and
OS Escherichia coli, O6.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Escherichia.
OX NCBI_TaxID=562, 217992;
[1]
RN SSOURCE FROM N.A.
RP MEDLINE=85155489; PubMed=6152241;
RA van Die I., Bergmans H.;
RT "Nucleotide sequence of the gene encoding the F72 fimbrial subunit of
RL a uropathogenic Escherichia coli strain.";
RN Gene 32:83-90(1984).
[2]
RN SSOURCE FROM N.A.
RP MEDLINE=92040048; PubMed=1682251;
RA Denich K., Blyn L.B., Craiu A., Braaten B.A., Hardy J., Low D.A.,
RA O'Hanley P.D.;
RT "DNA sequences of three papA genes from uropathogenic Escherichia
RL coli strains: evidence of structural and serological conservation.";
RN Infect. Immun. 59:3849-3858(1991).
[3]
RN SSOURCE FROM N.A.
RP STRAIN=06:HI / CFT073 / ATCC 700928;
RX MEDLINE=22388234; PubMed=12471157;
RA Welch R.A., Burland V., Plunkett G. III, Redford P., Roesch P.,
RA Rasko D., Buckles E.L., Liou S.-R., Boutin A., Hackett J., Stroud D.,
RA Mayhew G.F., Rose D.J., Zhou S., Schwartz D.C., Perna N.T.,
RA Mobley H.L.T., Donnenberg M.S., Blattner F.R.;
RT "Extensive mosaic structure revealed by the complete genome sequence
RL of uropathogenic Escherichia coli";
RN Proc. Natl. Acad. Sci. U.S.A. 99:17020-17024(2002).
CC -!- FUNCTION: Fimbriae (also called pili), polar filaments radiating
from the surface of the bacterium to a length of 0.5-1.5

```

```

CC micrometers and numbering 100-300 per cell, enable bacteria to
CC colonize the epithelium of specific host organs
CC -!- SIMILARITY: BELONGS TO THE FIMB/PAPA FAMILY OF FIMBRIA PROTEINS.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; M12861; AAA23778.1; -.
CC EMBL; M68060; AAA24278.1; -.
CC EMBL; AE016766; AANS2040.1; ALT_INIT.
CC PIR; A03496; YQECF2.
CC InterPro; IPR008966; Adhes bact.
CC InterPro; IPR000259; Fimbrial.
CC Pfam; PF00419; Fimbrial; 1.
KW Fimbria; Signal; Complete proteome.
FT SIGNAL 1 21
FT CHAIN 22 188 F7-2 FIMBRIAL PROTEIN.
FT DISULFID 43 82 PROBABLE.
SQ SEQUENCE 188 AA; 19184 MW; 0EEF750CFD843157 CRC64;

Query Match 10.9%; Score 84.5; DB 1; Length 188;
Best Local Similarity 25.0%; Pred. No. 1.4;
Matches 45; Conservative 26; Mismatches 76; Indels 33; Gaps 8;

QY 1 MKLLKVAFAAIVVS-GSALAGVVVPQWGGGNGGNG-----CGNSGPDSTLSIYQYGS 52
DB 2 IKSVIAGAVAVVVSFGAYAAPTIPOGQKVTNGTVDPAPGIDQAQADQSIDFGQV-- 59
QY 53 ANAALALQSDA---RKYD-QLVTRVVVTH-EMAHAGQADNSTIELTONGFRNATIDQW 106
DB 60 --SKLFLENDGESQPKSFDIKLINCIDITFKKAGGGAKGTIVLTFSGVPSGQSDML 117
QY 107 ---NAKNSDITVGOYGNNALVNQTASDSVM-----VRQVFGNNATANQY 151
DB 118 QTVGATNTAIVTDPHGKRVKFDGATATGVSILVDGNTIHFTAAVRKDGSGNFTVGAF 177

RESULT 15
ICEN_PANAN STANDARD; PRT; 1034 AA.
AC Q47879;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Ice nucleation protein inaU.
GN INAU.
OS Pantoea ananas (Erwinia uredovora).
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Pantoea.
OX NCBI_TaxID=553;
[1]
RN SSOURCE FROM N.A.
RP STRAIN=KUIN-3;
RX MEDLINE=94264407; PubMed=7764866;
RA Michigami Y., Watabe S., Abe K., Obata H., Arai S.;
RT "Cloning and sequencing of an ice nucleation active gene of Erwinia
RL uredovora.";
RN Biosci. Biotechnol. Biochem. 58:762-764(1994).
CC -!- FUNCTION: Ice nucleation proteins enable bacteria to nucleate
CC crystallization in supercooled water.
CC -!- SUBCELLULAR LOCATION: Outer membrane.
CC -!- DOMAIN: CONTAINS IMPERFECT REPEATS OF A CONSENSUS OCTAPEPTIDE
A-G-Y-G-S-T-X-T; FURTHER ON A 16-RESIDUE AND A REGIONAL 48-RESIDUE
PERIODICITY IS SUPERIMPOSED.
CC -!- MISCELLANEOUS: A STRUCTURAL MODEL IS SUGGESTED IN WHICH THE ICE
NUCLEATION PROTEIN DISPLAYS A SYMMETRY RELATED TO THAT OF ICE.
CC -!- SIMILARITY: Belongs to the bacterial ice nucleation protein
family.

```



```

CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; D14992; BAA03636.1; -
DR PIR; JC2143; JC2143.
DR HSP; P06620; IINA.
DR InterPro; IPR000258; Ice nucleatn.
DR Pfam; PF00818; Ice nucleation; 51.
DR PRINIS; PR00327; ICENUCLEATN.
DR PROSITE; PS00314; ICE NUCLEATION; 34.
KW Ice nucleation; Repeat; Outer membrane.
FT DOMAIN 162 993 OCTAPEPTIDE PERIODICITY.
SQ SEQUENCE 1034 AA; 103378 MW; FA222523D333EADD CRC64;

Query Match      10.9%; Score 84.5; DB 1; Length 1034;
Best Local Similarity 26.9%; Pred. No. 9.6;
Matches 45; Conservative 17; Mismatches 66; Indels 39; Gaps 9;

QY      12 IVVSGSALAGVWPQW--GGGNNHNGGNS-----GPDSTLSIYQYGSANAAL 57
Db      161 IATYGSTLSGTHQSQSLIAGYGTETAGDSSTLIAGYGTGTAGSDSTL-VAGYGSTQTAG 219

QY      58 ALQSDARKYDQLVTRV-----VTHEAHAGGADNSTIETONGFRNNATID----- 104
Db      220 EESSQAGYCGSTQTGMKGSDLTAGYGTGTAGDSSL-IAGYGTQTAGEDSSLTAGYGS 278

QY      105 -QWNAKNSDITVQYGGNNAALVNQTASDSSVMVRQVFGNNATNC 150
Db      279 TQTAQKGSDLTAG-YGSTGTA-----GADSSLI---AGYGSTQTAGE 316

```

Search completed: August 2, 2004, 14:49:30  
 Job time : 6.3 secs



GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: August 2, 2004, 14:39:17 ; Search time 29.7 Seconds  
(without alignments)  
1604.150 Million cell updates/sec

Title: US-09-543-407-20  
Perfect score: 774  
Sequence: 1 MKLLKVAFAAIVSGSALA.....DSSVMVRQVGFNNATANQY 151

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 1017041 seqs, 315518202 residues 1017041

Total number of hits satisfying chosen parameters:

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : SPTREMBL 25:\*  
1: sp\_archaea:\*  
2: sp\_bacteria:\*  
3: sp\_fungi:\*  
4: sp\_human:\*  
5: sp\_invertebrate:\*  
6: sp\_mammal:\*  
7: sp\_mhc:\*  
8: sp\_organelle:\*  
9: sp\_phase:\*  
10: sp\_plant:\*  
11: sp\_rodent:\*  
12: sp\_virus:\*  
13: sp\_vertebrate:\*  
14: sp\_unclassified:\*  
15: sp\_virus:\*  
16: sp\_bacteriaph:\*  
17: sp\_archaea:\*

Result No.	Score	Query Match	Length	DB ID	Description
1	680	87.9	152	2 O33802	Q33802 salmonella
2	582.5	75.3	150	2 Q7X243	Q7X243 citrobacter
3	552	71.3	149	2 Q7X240	Q7X240 citrobacter
4	503.5	65.1	152	16 Q8CW63	Q8CW63 escherichia
5	430.5	55.6	150	2 Q7X237	Q7X237 enterobacte
6	301	38.9	76	2 Q54059	Q54059 salmonella
7	122	15.8	29	2 Q9S3U5	Q9S3U5 escherichia
8	115	14.9	130	16 Q89J14	Q89J14 bradyrhizob
9	110	14.2	139	16 Q8E1H3	Q8E1H3 shewanella
10	106.5	13.8	502	16 Q8E1H4	Q8E1H4 shewanella
11	105.5	13.6	145	16 Q8U6N9	Q8U6N9 agrobacteri
12	105.5	13.6	153	16 Q89J16	Q89J16 bradyrhizob
13	98.99	12.8	573	10 Q9S3F2	Q9S3F2 arabidopsis
14	98.5	12.7	154	16 Q89J15	Q89J15 bradyrhizob
15	97.5	12.6	1765	16 Q7V8S5	Q7V8S5 prochloroco
16	97	12.5	157	16 Q88HG0	Q88HG0 pseudomonas

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	680	87.9	152	2 O33802	Q33802 salmonella
2	582.5	75.3	150	2 Q7X243	Q7X243 citrobacter
3	552	71.3	149	2 Q7X240	Q7X240 citrobacter
4	503.5	65.1	152	16 Q8CW63	Q8CW63 escherichia
5	430.5	55.6	150	2 Q7X237	Q7X237 enterobacte
6	301	38.9	76	2 Q54059	Q54059 salmonella
7	122	15.8	29	2 Q9S3U5	Q9S3U5 escherichia
8	115	14.9	130	16 Q89J14	Q89J14 bradyrhizob
9	110	14.2	139	16 Q8E1H3	Q8E1H3 shewanella
10	106.5	13.8	502	16 Q8E1H4	Q8E1H4 shewanella
11	105.5	13.6	145	16 Q8U6N9	Q8U6N9 agrobacteri
12	105.5	13.6	153	16 Q89J16	Q89J16 bradyrhizob
13	98.99	12.8	573	10 Q9S3F2	Q9S3F2 arabidopsis
14	98.5	12.7	154	16 Q89J15	Q89J15 bradyrhizob
15	97.5	12.6	1765	16 Q7V8S5	Q7V8S5 prochloroco
16	97	12.5	157	16 Q88HG0	Q88HG0 pseudomonas

Result No.	Score	Query Match	Length	DB ID	Description
1	680	87.9	152	2 O33802	Q33802 salmonella
2	582.5	75.3	150	2 Q7X243	Q7X243 citrobacter
3	552	71.3	149	2 Q7X240	Q7X240 citrobacter
4	503.5	65.1	152	16 Q8CW63	Q8CW63 escherichia
5	430.5	55.6	150	2 Q7X237	Q7X237 enterobacte
6	301	38.9	76	2 Q54059	Q54059 salmonella
7	122	15.8	29	2 Q9S3U5	Q9S3U5 escherichia
8	115	14.9	130	16 Q89J14	Q89J14 bradyrhizob
9	110	14.2	139	16 Q8E1H3	Q8E1H3 shewanella
10	106.5	13.8	502	16 Q8E1H4	Q8E1H4 shewanella
11	105.5	13.6	145	16 Q8U6N9	Q8U6N9 agrobacteri
12	105.5	13.6	153	16 Q89J16	Q89J16 bradyrhizob
13	98.99	12.8	573	10 Q9S3F2	Q9S3F2 arabidopsis
14	98.5	12.7	154	16 Q89J15	Q89J15 bradyrhizob
15	97.5	12.6	1765	16 Q7V8S5	Q7V8S5 prochloroco
16	97	12.5	157	16 Q88HG0	Q88HG0 pseudomonas

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	680	87.9	152	2 O33802	Q33802 salmonella
2	582.5	75.3	150	2 Q7X243	Q7X243 citrobacter
3	552	71.3	149	2 Q7X240	Q7X240 citrobacter
4	503.5	65.1	152	16 Q8CW63	Q8CW63 escherichia
5	430.5	55.6	150	2 Q7X237	Q7X237 enterobacte
6	301	38.9	76	2 Q54059	Q54059 salmonella
7	122	15.8	29	2 Q9S3U5	Q9S3U5 escherichia
8	115	14.9	130	16 Q89J14	Q89J14 bradyrhizob
9	110	14.2	139	16 Q8E1H3	Q8E1H3 shewanella
10	106.5	13.8	502	16 Q8E1H4	Q8E1H4 shewanella
11	105.5	13.6	145	16 Q8U6N9	Q8U6N9 agrobacteri
12	105.5	13.6	153	16 Q89J16	Q89J16 bradyrhizob
13	98.99	12.8	573	10 Q9S3F2	Q9S3F2 arabidopsis
14	98.5	12.7	154	16 Q89J15	Q89J15 bradyrhizob
15	97.5	12.6	1765	16 Q7V8S5	Q7V8S5 prochloroco
16	97	12.5	157	16 Q88HG0	Q88HG0 pseudomonas

Result No.	Score	Query Match	Length	DB ID	Description
1	680	87.9	152	2 O33802	Q33802 salmonella
2	582.5	75.3	150	2 Q7X243	Q7X243 citrobacter
3	552	71.3	149	2 Q7X240	Q7X240 citrobacter
4	503.5	65.1	152	16 Q8CW63	Q8CW63 escherichia
5	430.5	55.6	150	2 Q7X237	Q7X237 enterobacte
6	301	38.9	76	2 Q54059	Q54059 salmonella
7	122	15.8	29	2 Q9S3U5	Q9S3U5 escherichia
8	115	14.9	130	16 Q89J14	Q89J14 bradyrhizob
9	110	14.2	139	16 Q8E1H3	Q8E1H3 shewanella
10	106.5	13.8	502	16 Q8E1H4	Q8E1H4 shewanella
11	105.5	13.6	145	16 Q8U6N9	Q8U6N9 agrobacteri
12	105.5	13.6	153	16 Q89J16	Q89J16 bradyrhizob
13	98.99	12.8	573	10 Q9S3F2	Q9S3F2 arabidopsis
14	98.5	12.7	154	16 Q89J15	Q89J15 bradyrhizob
15	97.5	12.6	1765	16 Q7V8S5	Q7V8S5 prochloroco
16	97	12.5	157	16 Q88HG0	Q88HG0 pseudomonas

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	680	87.9	152	2 O33802	Q33802 salmonella
2	582.5	75.3	150	2 Q7X243	Q7X243 citrobacter
3	552	71.3	149	2 Q7X240	Q7X240 citrobacter
4	503.5	65.1	152	16 Q8CW63	Q8CW63 escherichia
5	430.5	55.6	150	2 Q7X237	Q7X237 enterobacte
6	301	38.9	76	2 Q54059	Q54059 salmonella
7	122	15.8	29	2 Q9S3U5	Q9S3U5 escherichia
8	115	14.9	130	16 Q89J14	Q89J14 bradyrhizob
9	110	14.2	139	16 Q8E1H3	Q8E1H3 shewanella
10	106.5	13.8	502	16 Q8E1H4	Q8E1H4 shewanella
11	105.5	13.6	145	16 Q8U6N9	Q8U6N9 agrobacteri
12	105.5	13.6	153	16 Q89J16	Q89J16 bradyrhizob
13	98.99	12.8	573	10 Q9S3F2	Q9S3F2 arabidopsis
14	98.5	12.7	154	16 Q89J15	Q89J15 bradyrhizob
15	97.5	12.6	1765	16 Q7V8S5	Q7V8S5 prochloroco
16	97	12.5	157	16 Q88HG0	Q88HG0 pseudomonas

Result No.	Score	Query Match	Length	DB ID	Description
1	680	87.9	152	2 O33802	Q33802 salmonella
2	582.5	75.3	150	2 Q7X243	Q7X243 citrobacter
3	552	71.3	149	2 Q7X240	Q7X240 citrobacter
4	503.5	65.1	152	16 Q8CW63	Q8CW63 escherichia
5	430.5	55.6	150	2 Q7X237	Q7X237 enterobacte
6	301	38.9	76	2 Q54059	Q54059 salmonella
7	122	15.8	29	2 Q9S3U5	Q9S3U5 escherichia
8	115	14.9	130	16 Q89J14	Q89J14 bradyrhizob
9	110	14.2	139	16 Q8E1H3	Q8E1H3 shewanella
10	106.5	13.8	502	16 Q8E1H4	Q8E1H4 shewanella
11	105.5	13.6	145	16 Q8U6N9	Q8U6N9 agrobacteri
12	105.5	13.6	153	16 Q89J16	Q89J16 bradyrhizob
13	98.99	12.8	573	10 Q9S3F2	Q9S3F2 arabidopsis
14	98.5	12.7	154	16 Q89J15	Q89J15 bradyrhizob
15	97.5	12.6	1765	16 Q7V8S5	Q7V8S5 prochloroco
16	97	12.5	157	16 Q88HG0	Q88HG0 pseudomonas

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	680	87.9	152	2 O33802	Q33802 salmonella
2	582.5	75.3	150	2 Q7X243	Q7X243 citrobacter
3	552	71.3	149	2 Q7X240	Q7X240 citrobacter
4	503.5	65.1	152	16 Q8CW63	Q8CW63 escherichia
5	430.5	55.6	150	2 Q7X237	Q7X237 enterobacte
6	301	38.9	76	2 Q54059	Q54059 salmonella
7	122	15.8	29	2 Q9S3U5	Q9S3U5 escherichia
8	115	14.9	130	16 Q89J14	Q89J14 bradyrhizob
9	110	14.2	139	16 Q8E1H3	Q8E1H3 shewanella
10	106.5	13.8	502	16 Q8E1H4	Q8E1H4 shewanella
11	105.5	13.6	145	16 Q8U6N9	Q8U6N9 agrobacteri
12	105.5	13.6	153	16 Q89J16	Q89J16 bradyrhizob
13	98.99	12.8	573	10 Q9S3F2	Q9S3F2 arabidopsis
14	98.5	12.7	154	16 Q89J15	Q89J15 bradyrhizob
15	97.5	12.6	1765	16 Q7V8S5	Q7V8S5 prochloroco
16	97	12.5	157	16 Q88HG0	Q88HG0 pseudomonas

Result No.	Score	Query Match	Length	DB ID	Description
1	680	87.9	152	2 O33802	Q33802 salmonella
2	582.5	75.3	150	2 Q7X243	Q7X243 citrobacter
3	552	71.3	149	2 Q7X240	Q7X240 citrobacter
4	503.5	65.1	152	16 Q8CW63	Q8CW63 escherichia
5	430.5	55.6	150	2 Q7X237	Q7X237 enterobacte
6	301	38.9	76	2 Q54059	Q54059 salmonella
7	122	15.8	29	2 Q9S3U5	Q9S3U5 escherichia
8	115	14.9	130	16 Q89J14	Q89J14 bradyrhizob
9	110	14.2	139	16 Q8E1H3	Q8E1H3 shewanella
10	106.5	13.8	502	16 Q8E1H4	Q8E1H4 shewanella
11	105.5	13.6	145	16 Q8U6N9	Q8U6N9 agrobacteri
12	105.5	13.6	153	16 Q89J16	Q89J16 bradyrhizob
13	98.99	12.8	573	10 Q9S3F2	Q9S3F2 arabidopsis
14	98.5	12.7	154	16 Q89J15	Q89J15 bradyrhizob
15	97.5	12.6	1765	16 Q7V8S5	Q7V8S5 prochloroco
16	97	12.5	157	16 Q88HG0	Q88HG0 pseudomonas

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	680	87.9	152	2 O33802	Q33802 salmonella
2	582.5	75.3	150	2 Q7X243	Q7X243 citrobacter
3	552	71.3	149	2 Q7X240	Q7X240 citrobacter
4	503.5	65.1	152	16 Q8CW63	Q8CW63 escherichia
5	430.5	55.6	150	2 Q7X237	Q7X237 enterobacte
6	301	38.9	76	2 Q54059	Q54059 salmonella
7	122	15.8	29	2 Q9S3U5	Q9S3U5 escherichia
8	115	14.9	130	16 Q89J14	Q89J14 bradyrhizob
9	110	14.2	139	16 Q8E1H3	Q8E1H3 shewanella
10	106.5	13.8	502	16 Q8E1H4	Q8E1H4 shewanella
11	105.5	13.6	145	16 Q8U6N9	Q8U6N9 agrobacteri
12	105.5	13.6	153	16 Q89J16	Q89J16 bradyrhizob
13	98.99	12.8	573	10 Q9S3F2	Q9S3F2 arabidopsis
14	98.5	12.7	154	16 Q89J15	Q89J15 bradyrhizob
15	97.5	12.6	1765	16 Q7V8S5	Q7V8S5 prochloroco
16	97	12.5	157	16 Q88HG0	Q88HG0 pseudomonas

Result No.	Score	Query Match	Length	DB ID	Description
1	680	87.9	152	2 O33802	Q33802 salmonella
2	582.5	75.3	150	2 Q7X243	Q7X243 citrobacter
3	552	71.3	149	2 Q7X240	Q7X240 citrobacter
4	503.5	65.1	152	16 Q8CW63	Q8CW63 escherichia
5	430.5	55.6	150	2 Q7X237	Q7X237 enterobacte
6	301	38.9	76	2 Q54059	Q54059 salmonella
7	122	15.8	29	2 Q9S3U5	Q9S3U5 escherichia
8	115	14.9	130	16 Q89J14	Q89J14 bradyrhizob
9	110	14.2	139	16 Q8E1H3	Q8E1H3 shewanella
10	106.5	13.8	502	16 Q8E1H4	Q8E1H4 shewanella
11	105.5	13.6	145	16 Q8U6N9	Q8U6N9 agrobacteri
12	105.5	13.6	153	16 Q89J16	Q89J1

```

Db      121 NNAALVNQTSASDSSVMVRQVGFNNATANOY 151
|||||
RESULT 2
Q7X243
ID Q7X243 PRELIMINARY; PRT; 150 AA.
AC Q7X243; 2003 (TrEMBLrel. 25, Created)
DT 01-OCT-2003 (TrEMBLrel. 25, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Curlin-csgA protein.
GN CSGA.
OS Citrobacter sp. Fec2.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Citrobacter.
OX NCBI_TaxID=213763;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Fec2.
RA Zogaj X., Bokranz W., Nimitz M., Romling U.;
RT "Production of Cellulose and Curli Fimbriae by Members of the Family
RT Enterobacteriaceae Isolated from the Human Gastrointestinal Tract.";
RL Infect. Immun. 72:4151-4158(2003).
DR EMBL; AJ515700; CAD56672.1; -.
SQ SEQUENCE 150 AA; 15016 MW; 1D7141B8D6973DC6 CRC64;

Query Match 75.3%; Score 582.5; DB 2; Length 150;
Best Local Similarity 78.8%; Pred. No. 2.3e-40;
Matches 119; Conservative 10; Mismatches 21; Indels 1; Gaps 1;

Qy      1 MLLKVAAPAAIVVSGSALAGVVPQWGGGNGHNGGNSGPPSTLSIYQYGSANAALALQ 60
|||||
Db      1 MLLQVAAPAAIVVSGSALAGVVPQWGGGNGHNGGNSGPPSTLSIYQYGSANAALALQ 59
|||||
Qy      61 SDARKYDQLVTRVVTHEMAHAGGADNSTIETQNGFRNATIDOWNAKNSDITVGOYGG 120
|||||
Db      60 SDARKSDTTIHQHGFGNGADVQGGSDNSTIDLTQNGFKNATIDOWNKGNKSDITVSOYGG 119
|||||
Qy      121 NNAALVNQTSASDSSVMVRQVGFNNATANOY 151
|||||
Db      120 HNAALVNQTSASDSSVLVHQVGFNNATANOY 150
|||||

RESULT 3
Q7X240
ID Q7X240 PRELIMINARY; PRT; 149 AA.
AC Q7X240; 2003 (TrEMBLrel. 25, Created)
DT 01-OCT-2003 (TrEMBLrel. 25, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Curlin-csgA protein.
GN CSGA.
OS Citrobacter freundii.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Citrobacter.
OX NCBI_TaxID=546;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Fec4;
RA Zogaj X., Bokranz W., Nimitz M., Romling U.;
RT "Production of Cellulose and Curli Fimbriae by Members of the Family
RT Enterobacteriaceae Isolated from the Human Gastrointestinal Tract.";
RL Infect. Immun. 72:4151-4158(2003).
DR EMBL; AJ515701; CAD56675.1; -.
SQ SEQUENCE 149 AA; 15260 MW; 946D52017F648FD CRC64;

Query Match 71.3%; Score 552; DB 2; Length 149;
Best Local Similarity 73.5%; Pred. No. 7.2e-38;
Matches 111; Conservative 18; Mismatches 20; Indels 2; Gaps 1;

Qy      1 MLLKVAAPAAIVVSGSALAGVVPQWGGGNGHNGGNSGPPSTLSIYQYGSANAALALQ 60
|||||

```

```

Db      1 MLLKVAAPAAIVVSGSALAGVVPQW--GGHHEGGGSGYSPDSSLSTIYQYGSNNANALQ 58
|||||
Qy      61 SDARKYDQLVTRVVTHEMAHAGGADNSTIETQNGFRNATIDOWNAKNSDITVGOYGG 120
|||||
Db      59 SDARKSDVTITQHGKNGAVVVGQGADDTSLKQTGFQNSATIDOWNAKNADISVTQFGG 118
|||||
Qy      121 NNAALVNQTSASDSSVMVRQVGFNNATANOY 151
|||||
Db      119 RAGALVNQTSASDSSVLVQVGFNNATANOY 149
|||||

RESULT 4
Q8CW63
ID Q8CW63 PRELIMINARY; PRT; 152 AA.
AC Q8CW63; 2003 (TrEMBLrel. 23, Created)
DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE Major curlin subunit precursor.
GN CSGA OR C1306.
OS Escherichia coli O6.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Escherichia.
OX NCBI_TaxID=217992;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=O6:H1 / CFT073 / ATCC 700928;
EX MEDLINE=22388234; PubMed=12471157;
RA Welch R.A., Burland V., Plunkett G. III, Redford P., Roesch P.,
RA Rasko D., Buckles E.L., Liou S.-R., Boutin A., Hackett J., Stroud D.,
RA Mayhew G.F., Rose D.J., Zhou S., Schwartz D.C., Perna N.T.,
RA Mobley H.L.T., Donnenberg M.S., Blattner F.R.;
RT "Extensive mosaic structure revealed by the complete genome sequence
RT of uropathogenic Escherichia coli.";
RL Proc. Natl. Acad. Sci. U.S.A. 99:17020-17024(2002).
DR EMBL; AE016759; AN79779.1; -.
KW Complete proteome.
SQ SEQUENCE 152 AA; 15064 MW; 3BA57F34C1240E83 CRC64;

Query Match 65.1%; Score 503.5; DB 16; Length 152;
Best Local Similarity 67.1%; Pred. No. 7.1e-34;
Matches 102; Conservative 20; Mismatches 29; Indels 1; Gaps 1;

Qy      1 MLLKVAAPAAIVVSGSALAGVVPQW--GGGNGHNGGNSGPPSTLSIYQYGSANAALAL 59
|||||
Db      1 MLLKVAAPAAIVVSGSALAGVVPQYGGGGNGHNGGNSGPPSTLSIYQYGGNSALAQ 60
|||||
Qy      60 QSDARKYDQLVTRVVTHEMAHAGGADNSTIETQNGFRNATIDOWNAKNSDITVGOYGG 119
|||||
Db      61 QADARNSDLTITQHGKNGADVQGGSDSDSIDLTQNGFNSATIDOWNKDKSTMTVKQFG 120
|||||
Qy      120 GNAALVNQTSASDSSVMVRQVGFNNATANOY 151
|||||
Db      121 GNGAAVDQTSASNSVNVTVQVGFNNATANOY 152
|||||

RESULT 5
Q7X237
ID Q7X237 PRELIMINARY; PRT; 150 AA.
AC Q7X237; 2003 (TrEMBLrel. 25, Created)
DT 01-OCT-2003 (TrEMBLrel. 25, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Curlin-csgA protein.
GN CSGA.
OS Enterobacter sakazakii.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Enterobacter.
OX NCBI_TaxID=28141;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Fec39;
RA Zogaj X., Bokranz W., Nimitz M., Romling U.;

```

```

RT "Production of Cellulose and Curli Fimbriae by Members of the Family
RL Enterobacteriaceae Isolated from the Human Gastrointestinal Tract.";
RL Infect. Immun. 72:4151-4158 (2003).
DR EMBL: AJ515702; CAD56678.1; -. 5DBBB2D872DF15F3 CRC64;
SQ SEQUENCE 150 AA; 15112 MW; 5DBBB2D872DF15F3 CRC64;

Query Match 55.6%; Score 430.5; DB 2; Length 150;
Best Local Similarity 58.9%; Pred. No. 6.9e-28;
Matches 89; Conservative 27; Mismatches 34; Indels 1; Gaps 1;

QY 1 MKLLKVAAPAAIVVSGSALAGVVPQWGGNGGNGSSGPDSTLSIYQGSANAALALQ 60
DB |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1 MKFIKVAALAAIVVSGSAGVAGNIG--GGWGHGGHGGYGGPNSLTNIYONGGNGSALALQ 59

QY 61 SDARKYDQLVTRVVTTHMAHAGQADNSTIETQNGFRNNATIDOWNAKNSDITYQYGG 120
DB |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
60 TDARNVLNISQTGGGNGADVGGSDSSINTQNGFGNSATLDQNSKDSVMVNSQYGG 119

QY 121 NNAALVNOTASDSSVMVRQVFGNNATANQY 151
DB |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
120 LNCALVDQTASNSTVNTVQTGFGNHATAHQY 150

RESULT 6
Q54069 PRELIMINARY; PRT; 76 AA.
ID Q54069
AC Q54069
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE SEF17 fimbria (fragment).
GN AGFA.
OS Salmonella enteritidis.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Salmonella.
OX NCBI_TaxID=592;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=SE30;
RA Cox J.M., Eglezos S., Woolcock J.B.;
RT "Virulence of Salmonella enteritidis in chickens correlates with
RT colony morphology and expression of SEF17 fimbriae.";
RL Submitted (APR-1996) to the EMBL/GenBank/DBJ databases.
DR EMBL: U53207; AAA98671.1; -.
FT NON TER 1
FT NON TER 76
SQ SEQUENCE 76 AA; 7704 MW; 2FD5411241A7BCB1 CRC64;

Query Match 38.9%; Score 301; DB 2; Length 76;
Best Local Similarity 80.3%; Pred. No. 1.3e-17;
Matches 6; Conservative 3; Mismatches 12; Indels 0; Gaps 0;

QY 30 GNNHGGNGSGPDSTLSIYQGSANAALALQSDARKYDQLVTRVVTTHMAHAGQADNST 89
DB |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1 GNNHGGNGSGPDSTLSIYQGSANAALALQSDARKSETTITQSGYNGADVGGADNST 60

QY 90 IELTQNGFRNNATIDQ 105
DB |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
61 IELTQNGFRNNATIDQ 76

RESULT 7
Q9S3J5 PRELIMINARY; PRT; 29 AA.
ID Q9S3J5
AC Q9S3J5
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT 01-MAR-2001 (TrEMBLrel. 16, Last annotation update)
DE Curlin subunit monomer (Fragment).
GN CSCA.
OS Escherichia coli.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Escherichia.

```

```

OX NCBI_TaxID=562;
RN [1]
RP SEQUENCE FROM N.A.
RC TRANSPOSON=Insertion sequence IS1;
RX MEDLINE=99314153; PubMed=10386375;
RA La Ragione R.M., Collighan R.J., Woodward M.J.;
RT "Non-curliation of Escherichia coli O78:K80 isolates associated with
RT IS1 inserti on in csGB and reduced persistence in poultry infection.";
RL FEMS Microbiol. Lett. 175:247-253 (1999).
DR EMBL: AJ131756; CAB45380.1; -.
FT NON TER 29
FT NON TER 29
SQ SEQUENCE 29 AA; 2789 MW; E290DFC07ABBB243 CRC64;

Query Match 15.8%; Score 122; DB 2; Length 29;
Best Local Similarity 89.7%; Pred. No. 0.0019;
Matches 26; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 MKLLKVAAPAAIVVSGSALAGVVPQWGGG 29
DB |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1 MKLLKVAAPAAIVVSGSALAGVVPQYGGG 29

RESULT 8
Q89JI4 PRELIMINARY; PRT; 130 AA.
ID Q89JI4
AC Q89JI4
DT 01-JUN-2003 (TrEMBLrel. 24, Created)
DT 01-JUN-2003 (TrEMBLrel. 24, Last sequence update)
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE BL15299 protein.
GN BL15299.
OS Bradyrhizobium japonicum.
OC Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;
OC Bradyrhizobiaceae; Bradyrhizobium.
OX NCBI_TaxID=375;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=USDA 110;
RX MEDLINE=2248498; PubMed=12597275;
RA Kaneko T., Nakamura Y., Sato S., Minamisawa K., Uchiumi T.,
RA Sasamoto S., Watanabe A., Ideawa K., Iiiguchi M., Kawashima K.,
RA Kohara M., Matsumoto M., Shimpo S., Tsuruoka H., Wada T., Yamada M.,
RA Tabata S.;
RT "Complete genomic sequence of nitrogen-fixing symbiotic bacterium
RT Bradyrhizobium japonicum USDA110.";
RL DNA Res. 9:189-197 (2002).
DR EMBL: AP005954; BAC50564.1; -.
KW Complete proteome.
SQ SEQUENCE 130 AA; 12699 MW; ACFB2D6A48D260F CRC64;

Query Match 14.9%; Score 115; DB 16; Length 130;
Best Local Similarity 25.9%; Pred. No. 0.046;
Matches 41; Conservative 27; Mismatches 48; Indels 42; Gaps 6;

QY 4 LKVAAPAAIVVSGSALAGVVPQWGGG-----GNHNGG-----GNSSGPDSTLSIYQGS 52
DB |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1 MEITYLVATAIALSALTITVDAQAGNSASVLOFGITNSSFISQTGSTNNATL---QFGA 57

QY 53 ANAALALQSDARKYDQLVTRVVTTHMAHAGQADNSTIETQNGFRNNATIDOWNAKNSD 112
DB |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
58 TWTATLQTGS-----LLTVNTAVTGGGTTAT-----ASNTA 90

QY 113 ITVGQYGGNNAALVNOTASDSSVMVRQVFGNNATANQ 150
DB |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
91 LT-GQVGGSSNSLIQIGANNNTAGVQGLGILNGSTILQ 127

RESULT 9
Q8EIH3 PRELIMINARY; PRT; 139 AA.
ID Q8EIH3
AC Q8EIH3
DT 01-MAR-2003 (TrEMBLrel. 23, Created)
DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)

```

DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)  
 DE Minor curlin subunit CsgB, putative.  
 GN SO0866.  
 OS Shewanella oneidensis.  
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Alteromonadales;  
 OC Alteromonadaceae; Shewanella.  
 OX NCBI\_TaxID=70863;  
 [1]  
 RN SEQUENCE FROM N.A.  
 RP STRAIN=MR-1;  
 RC MEDLINE=22297686; PubMed=12368813;  
 RA Heidelberg J.F., Paulsen I.T., Nelson K.E., Gaidos E.J., Nelson W.C.,  
 RA Read T.D., Eisen J.A., Seshadri R., Ward N., Methe B., Clayton R.A.,  
 RA Meyer T., Tsapin A., Scott J., Beanan M., Brinkac L., Daugherty S.,  
 RA DeBoy R.T., Dodson R.J., Durkin A.S., Haft D.H., Kolonay J.F.,  
 RA Madupu R., Peterson J.D., Unayam L.A., White O., Wolf A.M.,  
 RA Vamathevan J., Weidman J., Imprim M., Lee K., Berry K., Lee C.,  
 RA Mueller J., Khouri H., Gill J., Utterback T.R., McDonald L.A.,  
 RA Feldblyum T.V., Smith H.O., Venter J.C., Nealon K.H., Fraser C.M.,  
 RA "Genome sequence of the dissimilatory metal ion-reducing bacterium  
 RT Shewanella oneidensis."  
 RL Nat. Biotechnol. 20:1118-1123(2002).  
 DR EMBL; AE015532; AAN53942.1; -  
 DR TIGR; SO0866; -  
 KW Complete proteome.  
 SQ SEQUENCE 139 AA; 14811 MW; 41EC1FA76957920 CRC64;  
 Query Match 14.2%; Score 110; DB 16; Length 139;  
 Best Local Similarity 28.3%; Pred. No. 0.13;  
 Matches 30; Conservative 18; Mismatches 34; Indels 24; Gaps 3;  
 QY 39 SGPSTLSIYQVGSANAALQSDARKYDQLVTRVVTHEMAHAGQADNSTIELTQNGFR 98  
 DB 41 SGRDLIDLVQQTANQGVFQS-----GSDNSAY-VTQAGND 77  
 QY 99 NNATIDQWNAKNSDITVQYGGNNAALVNQTSADSSVMVQVGFNGFN 144  
 DB 78 NISLVITQIGT-NNEVQLLVQGAQNKASITQIGNDNLVQLNQLSGN 122  
 RESULT 10  
 Q8EIH4 PRELIMINARY; PRT; 502 AA.  
 AC Q8EIH4;  
 DT 01-MAR-2003 (TrEMBLrel. 23, Created)  
 DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)  
 DE Conserved hypothetical protein.  
 GN SO0865.  
 OS Shewanella oneidensis.  
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Alteromonadales;  
 OC Alteromonadaceae; Shewanella.  
 OX NCBI\_TaxID=70863;  
 [1]  
 RN SEQUENCE FROM N.A.  
 RP STRAIN=MR-1;  
 RC MEDLINE=22297686; PubMed=12368813;  
 RA Heidelberg J.F., Paulsen I.T., Nelson K.E., Gaidos E.J., Nelson W.C.,  
 RA Read T.D., Eisen J.A., Seshadri R., Ward N., Methe B., Clayton R.A.,  
 RA Meyer T., Tsapin A., Scott J., Beanan M., Brinkac L., Daugherty S.,  
 RA DeBoy R.T., Dodson R.J., Durkin A.S., Haft D.H., Kolonay J.F.,  
 RA Madupu R., Peterson J.D., Unayam L.A., White O., Wolf A.M.,  
 RA Vamathevan J., Weidman J., Imprim M., Lee K., Berry K., Lee C.,  
 RA Mueller J., Khouri H., Gill J., Utterback T.R., McDonald L.A.,  
 RA Feldblyum T.V., Smith H.O., Venter J.C., Nealon K.H., Fraser C.M.,  
 RA "Genome sequence of the dissimilatory metal ion-reducing bacterium  
 RT Shewanella oneidensis."  
 RL Nat. Biotechnol. 20:1118-1123(2002).  
 DR EMBL; AE015532; AAN53941.1; -  
 DR TIGR; SO0865; -  
 KW Hypothetical protein; Complete proteome.  
 SQ SEQUENCE 502 AA; 52441 MW; D08CA23D6C46B62D CRC64;

Query Match 13.8%; Score 106.5; DB 16; Length 502;  
 Best Local Similarity 23.1%; Pred. No. 1.2;  
 Matches 40; Conservative 30; Mismatches 72; Indels 31; Gaps 5;  
 QY 10 AAIIVVSGSALAGVVPQWGG-----GNHNG-----GNS-----SGPD 42  
 DB 163 AVFRVEGDNDGDIKQYGNNGNAGLIAIDLISANVGNNDVSVQIINNFGAAKGIAGND 222  
 QY 43 STLSIYQVGSANAALQSDARKYDQLVTRVVTHEMAHAGQ-CADNSTIELTQNGFRN-- 99  
 DB 223 NSVDIYQKGNHTGTFYVALAGSENDISMEQEGSNNTAYLSMTTGGDNTVDITQDGSNTV 282  
 QY 100 -NATIDQWNAKNSDITVQYGGNNAALVNQTSADSSVMVQVGFNGNATANQY 151  
 DB 283 GDSLIADIQGDNDITIKQKDSNGAEFFQVWGSDNSVDLQKRGDANFATFGAY 335  
 RESULT 11  
 Q8U6N9 PRELIMINARY; PRT; 145 AA.  
 AC Q8U6N9;  
 DT 01-JUN-2002 (TrEMBLrel. 21, Created)  
 DT 01-JUN-2002 (TrEMBLrel. 21, Last sequence update)  
 DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)  
 DE Hypothetical protein Atu4768.  
 GN ATU4768 OR AGR L 228.  
 OS Agrobacterium tumefaciens (strain C58 / ATCC 33970).  
 OC Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;  
 OC Rhizobiaceae; Rhizobium/Agrobacterium group; Agrobacterium.  
 OX NCBI\_TaxID=176299;  
 [1]  
 RN SEQUENCE FROM N.A.  
 RP MEDLINE=21608550; PubMed=11743193;  
 RA Wood D.W., Stuhel J.C., Kaul R., Monks D.E., Kitajima J.P.,  
 RA Okura Y.K., Zhou Y., Chen L., Wood G.E., Almeida N.F. Jr., Woo L.,  
 RA Chen Y., Paulsen I.T., Eisen J.A., Karp P.D., Bovee D. Sr.,  
 RA Chapman P., Clendenning J., Deatherage G., Gillet W., Grant C.,  
 RA Kutyavin T., Levy R., Li M.-J., McClelland E., Palmeri A.,  
 RA Raymond C., Rouse G., Saenphimmachak C., Wu Z., Romero P., Gordon D.,  
 RA Zhang S., Yoo H., Tao Y., Biddle P., Jung M., Krespan W., Perry M.,  
 RA Gordon-Kamm B., Liao L., Kim S., Hendrick C., Zhao Z.-Y., Dolan M.,  
 RA Chumley F., Tingey S.V., Tomb J.-F., Gordon M.P., Olson M.V.,  
 RA Nester E.W.;  
 RA "The genome of the natural genetic engineer Agrobacterium tumefaciens  
 RT C58."  
 RL Science 294:2317-2323(2001).  
 [2]  
 RN SEQUENCE FROM N.A.  
 RP MEDLINE=21608551; PubMed=11743194;  
 RA Goodner B., Hinkie G., Gattung S., Miller N., Blanchard M.,  
 RA Qurollo B., Goldman B.S., Cao Y., Askenazi M., Halling C., Mullin L.,  
 RA Houmieu K., Gordon J., Vaudin M., Tarchouk O., Epp A., Liu F.,  
 RA Wollam C., Allinger M., Doughty J., Scott C., Lappas C., Markelz B.,  
 RA Flanagan C., Crowell C., Gurson J., Lomo C., Sear C., Strub G.,  
 RA Cielo C., Slater S.;  
 RA "Genome sequence of the plant pathogen and biotechnology agent  
 RT Agrobacterium tumefaciens C58."  
 RL Science 294:2323-2328(2001).  
 DR EMBL; AE009405; AAL45562.1; -  
 DR EMBL; AE008209; AAK88682.1; -  
 DR PIR; AD3143; AD3143.  
 DR PIR; H98144; H98144.  
 KW Hypothetical protein; Complete proteome.  
 SQ SEQUENCE 145 AA; 14984 MW; DEDC870E1713D51A CRC64;  
 Query Match 13.6%; Score 105.5; DB 16; Length 145;  
 Best Local Similarity 23.2%; Pred. No. 0.32;  
 Matches 36; Conservative 27; Mismatches 61; Indels 31; Gaps 4;  
 QY 3 LLKVAAFAAIVVSGSALAGVVPQWGG-----GNHNGGSSGSPDSTLSIYQVGSANA 55  
 DB 1 MIRKSFASALVALVUSAAAPANDVRIEQVWSNAGAGQEGNRIYQNGYNR 60

QY 56 ALALQSDARKYDQVTRVVTTHMAHAGQADNSTIETONGFRNATIDQNAKNSDITV 115  
 Db 61 IVGHQ-----YGR-----HNSAVGQGHNDYSGTITQNGRNVAGI----- 96  
 QY 116 GQYGGNNAALVNOTASDSSVMVQVGFNNATANQ 150  
 Db 97 GQFGSHTTILTDQNGNIAAGVQVGRGCSANVSQ 131

RESULT 12  
 Q89JI6 PRELIMINARY; PRT; 153 AA.  
 AC Q89JI6; 24, Created)  
 DT 01-JUN-2003 (TREMELrel. 24, Last sequence update)  
 DT 01-JUN-2003 (TREMELrel. 24, Last sequence update)  
 DE CsgB protein.  
 GN CsgB OR BL5297.  
 OS Bradyrhizobium japonicum.  
 OC Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;  
 OC Bradyrhizobiaceae; Bradyrhizobium.  
 NCBI\_TaxID=375;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=USDA 110;  
 RX MEDLINE=22484958; PubMed=12597275;  
 RA Kaneko T., Nakamura Y., Sato S., Minamisawa K., Uchiyama T.,  
 RA Sasamoto S., Watanabe A., Iidesawa K., Iriguchi M., Kawashima K.,  
 RA Kohara M., Matsumoto M., Shimo S., Teirukawa H., Wada T., Yamada M.,  
 RA Tabata S.;  
 RT "Complete genomic sequence of nitrogen-fixing symbiotic bacterium  
 RT Bradyrhizobium japonicum USDA110.";  
 RL DNA Res 9:189-197(2002).  
 DR ENBL; AF005954; BAC50562.1; -  
 KW Complete proteome.  
 SQ SEQUENCE 153 AA; 15991 MW; 4CE7LDEAC375145B CRC64;

Query Match 13.6%; Score 105.5; DB 16; Length 153;  
 Best Local Similarity 25.7%; Pred. No. 0.34;  
 Matches 39; Conservative 35; Mismatches 55; Indels 23; Gaps 7;

QY 2 KLKVA-AFAIVVSGSALAGVVPQGGGNGHNGSGSPDSTL-SIYQGSANAALAL 59  
 Db 10 RVLAVALLAAGATQASAGSIQ-----SVTPNVSIETIVQFGNDVQPVTI 58  
 QY 60 QSDARKYDQVTRVVTTHMAHAGQADNSTIETONGFRNATIDQNAKNSDITVQY 119  
 Db 59 EENSRLN---IARV-----QIGSGTVDATL--IQNGTRNYANVIQGG-TTNAVQSG 108  
 QY 120 GNAALVNOTASDSSVMVQVGFNNATANQY 151  
 Db 109 LSNATDITQGNSTNALLQIGDMNSGAVRQF 140

RESULT 13  
 Q89SAF2 PRELIMINARY; PRT; 573 AA.  
 AC Q89SAF2;  
 DT 01-MAY-2000 (TREMELrel. 13, Created)  
 DT 01-MAY-2000 (TREMELrel. 13, Last sequence update)  
 DT 01-OCT-2003 (TREMELrel. 25, Last annotation update)  
 DE F3F19.21 protein (Hypochemical protein).  
 GN F3F19.21 OR AtG13190, F3F19.21 OR AtG13190.  
 OS Arabidopsis thaliana (Mouse-ear cress).  
 OC Eukaryota; Viridiplantae; Streptophyta; Tracheophyta;  
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids;  
 OC eurosids II; Brassicales; Brassicaceae; Arabidopsis.  
 NCBI\_TaxID=3702;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=cv. Columbia;  
 RA Vysotskaia V.S., Schwartz J.R., Yu G., Toriumi M., Lee J.M., Lenz C.,  
 RA Liu S., Li J., Kremenetskaia I., Luros J., Ngan I., Gonzalez A.,

RA Altafi H., Araujo R., Chao Q., Conn L., Conway A.B., Dunn P.,  
 RA Hansen N., Huizar L., Kim C., Palm C., Rowley D., Shinn P., Walker M.,  
 RA Davis R.W., Ecker J.R., Federspiel N.A., Theologis A.;  
 RT "Arabidopsis thaliana chromosome 1 BAC F3F19 sequence.";  
 RL Submitted (APR-1999) to the EMBL/GenBank/DBJ databases.  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=cv. Columbia;  
 RA Theologis A.;  
 RL Submitted (MAY-1999) to the EMBL/GenBank/DBJ databases.  
 RN [3]  
 RP SEQUENCE FROM N.A.  
 RA Southwick A., Karlin-Neumann G., Nguyen M., Lam B., Miranda M.,  
 RA Palm C.J., Bowser L., Jones T., Banh J., Carninci P., Chen H.,  
 RA Cheuk R., Chung M.K., Hayashizaki Y., Ishida J., Kamiya A., Kawai J.,  
 RA Kim C., Lin J., Liu S.X., Narusaka M., Pham P.K., Sakano H.,  
 RA Sakurai T., Satou M., Seki M., Shinn P., Yamada K., Shinozaki K.,  
 RA Ecker J., Theologis A., Davis R.W.;  
 RL Submitted (NOV-2001) to the EMBL/GenBank/DBJ databases.  
 RN [4]  
 RP SEQUENCE FROM N.A.  
 RA Nguyen M., Southwick A., Tripp M., Palm C.J., Jones T., Wu T.,  
 RA Carninci P., Chen H., Cheuk R., Chan M.M., Chang C.H., Dale J.M.,  
 RA Deng J.M., Hayashizaki Y., Hsuan V.W., Lee J.M., Ishida J., Kamiya A.,  
 RA Kawai J., Kim C.J., Narusaka M., Onodera C.S., Quach H.L., Sakurai T.,  
 RA Satou M., Seki M., Shinn P., Tang C.C., Toriumi M., Wong C., Wu H.C.,  
 RA Yamada K., Yu G., Yuan S., Shinozaki K., Ecker J., Theologis A.,  
 RA Davis R.W.;  
 RL Submitted (DSC-2002) to the EMBL/GenBank/DBJ databases.  
 DR ENBL; AC007357; AAD31072.1; -  
 DR ENBL; AY062527; AAL32605.1; -  
 DR ENBL; BT002575; AAO00935.1; -  
 DR PIR; C86266; C86266.  
 DR GO; GO:0003676; F:nucleic acid binding; IEA.  
 DR InterPro; IPR00504; RNA\_rec\_mot.  
 DR Pfam; PF00076; rrm; 1.  
 DR SMART; SM00360; RRM; 1.  
 DR PROSITE; PS0102; RRM; 1.  
 DR Hypothetical protein.  
 KW SEQUENCE 573 AA; 61400 MW; 0D1C10384B86C2BC CRC64;

Query Match 12.8%; Score 99; DB 10; Length 573;  
 Best Local Similarity 25.4%; Pred. No. 5.9;  
 Matches 34; Conservative 24; Mismatches 46; Indels 30; Gaps 6;

QY 13 VVSGSALAGVVPQGGGNGHNGSGSPDSTL-SIYQGSANAALALQSDARKYDQVTR 72  
 Db 84 IVSGGTVEG---KYRNDGCHNG---ISGPTRSVPYQASSFGAKGLNID----- 127  
 QY 73 VVTHMAHAGQADNSTIETONGFRNNA-TIDOWNAKNS-----DITVQYGGNN 122  
 Db 128 IQSNKIAQGG---STTVLNNHGFSGNAVNPMPVHNSYGAPOGACQIPVSGMSVNP 183  
 QY 123 AALVNQATASDSSVM 136  
 Db 184 NVMMNKSPQTQSFV 197

RESULT 14  
 Q89JI5 PRELIMINARY; PRT; 154 AA.  
 AC Q89JI5;  
 DT 01-JUN-2003 (TREMELrel. 24, Created)  
 DT 01-JUN-2003 (TREMELrel. 24, Last sequence update)  
 DT 01-JUN-2003 (TREMELrel. 24, Last annotation update)  
 DE B15298 protein.  
 GN B15298.  
 OS Bradyrhizobium japonicum.  
 OC Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;  
 OC Bradyrhizobiaceae; Bradyrhizobium.  
 NCBI\_TaxID=375;  
 RN [1]  
 RP SEQUENCE FROM N.A.

Search completed: August 2, 2004, 14:54:39  
Job time : 30.7 secs



GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: August 2, 2004, 14:35:42 ; Search time 44.9 Seconds  
(without alignments)  
950.215 Million cell updates/sec

Title: US-09-543-407-22

Perfect score: 776  
Sequence: 1 MKLLKVAFAIAIVVSGSALA.....DSSVMVRQVFGNNATANY 151

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 1586107 seqs, 282547505 residues

Total number of hits satisfying chosen parameters: 1586107

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : A\_Geneseq\_29Jbn04.\*

- 1: Geneseq1980s.\*
- 2: Geneseq1990s.\*
- 3: Geneseq2000s.\*
- 4: Geneseq2001s.\*
- 5: Geneseq2002s.\*
- 6: Geneseq2003as.\*
- 7: Geneseq2003bs.\*
- 8: Geneseq2004s.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	776	100.0	151	3	AAB36351
2	712	91.8	151	3	AAB36354
3	682	87.9	151	2	AAR74625
4	682	87.9	151	3	AAB36341
5	677	87.2	151	2	AAW23570
6	657	84.7	151	3	AAB36355
7	619	79.8	151	3	AAB36350
8	609	78.5	151	3	AAB36352
9	604	77.8	151	3	AAB36346
10	602	77.6	151	3	AAB36347
11	601	77.4	151	3	AAB36353
12	600	77.3	151	3	AAB36349
13	567	73.1	151	3	AAB36348
14	521	67.1	151	3	AAB36343
15	516	66.5	151	7	ABR82651
16	497	64.0	120	2	AAW23761
17	497	64.0	120	2	AAW23569
18	443	57.1	142	2	AAW23664
19	371	47.8	122	2	AAW23663
20	146	18.8	45	3	AAB36316
21	132	17.0	22	3	AAB36318
22	115	14.8	22	3	AAB36325
23	115	14.8	22	3	AAB36339
24	115	14.8	22	3	AAB36320
25	113	14.6	24	7	ABR82644

26	111	14.3	22	3	AAB36322	Aab36322 Salmonell
27	111	14.3	22	3	AAB36327	Aab36327 Salmonell
28	111	14.3	22	3	AAB36337	Aab36337 Salmonell
29	109	14.0	23	3	AAB36340	Aab36340 Salmonell
30	109	14.0	23	3	AAB36324	Aab36324 Salmonell
31	109	14.0	23	3	AAB36319	Aab36319 Salmonell
32	102	13.1	26	7	ABR82645	ABR82645 E. coli v
33	98	12.6	26	7	ABR82645	ABR82645 E. coli c
34	97.5	12.6	520	6	AAO16497	AAO16497 Argiope t
35	96	12.4	19	3	AAB36323	Aab36323 Salmonell
36	96	12.4	19	3	AAB36336	Aab36336 Salmonell
37	96	12.4	19	3	AAB36328	Aab36328 Salmonell
38	96	12.4	2309	4	ABE6232	ABE6232 Drosophil
39	95.5	12.3	252	8	ADE83885	Ade83885 Chemokine
40	95	12.2	597	4	AAU08231	AAU08231 Polypept
41	94.5	12.2	738	2	AAW56163	AAW56163 New DNA s
42	92.5	11.9	151	3	AAB36344	Aab36344 Escherich
43	92.5	11.9	251	5	ABP45119	ABP45119 Human Bly
44	91	11.7	1397	7	ADD42761	ADD42761 Chlamydia
45	91	11.7	1751	5	ABG91039	ABG91039 Chlamydia

## ALIGNMENTS

## RESULT 1

AAB36351  
ID AAB36351 standard; protein; 151 AA.  
AC AAB36351;  
XX  
DT 26-FEB-2001 (first entry)  
XX  
DE Agfa::PT3#6 amino acid sequence SEQ ID NO:22.

XX Salmonella; agfa; chromosomal gene replacement; fimbria; epitope;  
KW vaccine; immune response; immunogen.  
XX

OS Salmonella enteritidis.

OS Escherichia coli.

OS Synthetic.

XX WO200060102-A2.

XX

PD 12-OCT-2000.

XX

PF 05-APR-2000; 2000WO-CA000356.

XX

PR 05-APR-1999; 99US-0127888P.

XX (UYVI-) UNIV VICTORIA.

PA

XX

PI White AP, Doran JL, Collison SK, Kay WW;

XX WPI; 2000-672631/65.

DR N-PSDB; AAC64627.

XX

PT Recombinant agfa gene having a segment replaced by a foreign DNA sequence  
PT which encodes foreign epitope or antigen, expresses recombinant Agfa  
PT protein useful for eliciting immune response in animal.

XX

PS Disclosure; Page 137; 139pp; English.

XX

CC The present invention describes a recombinant agfa gene (1) where a  
CC segment of the gene has been replaced by a segment of a foreign DNA  
CC sequence which encodes a foreign epitope or antigen. Also described are:  
CC (1) use of thin aggregative fimbriae (SEPI7/TAI) nucleation depended  
CC assembly system of strains of Salmonella, Escherichia coli and  
CC Enterobacteriaceae for the production of fimbriae comprising recombinant  
CC Agfa, CsgA and Agfa-homologue fimbria subunits, respectively; (2)  
CC directing recombination of a recombinant gene into the chromosome of the  
CC homologous species; (3) directing recombination of a recombinant gene  
CC back into the chromosome of the homologous species, replacing the native

CC copy of that gene; and (4) eliciting an immune response in an animal,  
 CC comprising separating an amino acid polymer comprising a recombinant Agfa  
 CC protein containing a replacement segment or segments of foreign amino  
 CC acid sequence or sequences grown on a Salmonella, E. coli or  
 CC Enterobacteriaceae host cell, from the host cell and introducing the  
 CC polymer into the animal in conjunction with a carrier or diluent. (1) is  
 CC useful for the expression of recombinant Agfa protein which is useful for  
 CC eliciting an immune response in an animal. In a fimbrial presentation  
 CC system the heterologous antigens are presented in high numbers (up to  
 CC 500,000 copies/cell), the hybrid fimbrial protein possesses both the  
 CC immunogenicity and adhesion properties relevant for an efficient live  
 CC vaccine, the carrier fimbrial subunit proteins are usually strong  
 CC immunogens, which may be important for directing an immune response  
 CC against the inserted epitope, and hybrid fimbriae are easy and  
 CC inexpensive to purify in large amount. The present sequence is given in  
 CC the exemplification of the present invention  
 XX  
 SQ Sequence 151 AA;

Query Match 100.0%; Score 776; DB 3; Length 151;  
 Best Local Similarity 100.0%; Pred. No. 9.5e-68; Indels 0; Gaps 0;  
 Matches 151; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 MLLKLVAAFAAIVVSGSALAGVVPQWGGGNGHNGGNSGPDSTLSIYQYGSANAALALQ 60  
 DB 1 MLLKLVAAFAAIVVSGSALAGVVPQWGGGNGHNGGNSGPDSTLSIYQYGSANAALALQ 60  
 QY 61 SDARKSETTITQSGYNGADVGQADNYDOLVTRVVTHEMAHADQWNAKNSDITVGYGG 120  
 DB 61 SDARKSETTITQSGYNGADVGQADNYDOLVTRVVTHEMAHADQWNAKNSDITVGYGG 120  
 QY 121 NNAALVNTASDSSVMVRQVGFNNATANQY 151  
 DB 121 NNAALVNTASDSSVMVRQVGFNNATANQY 151

RESULT 2  
 AAB36354  
 ID AAB36354 standard; protein; 151 AA.  
 AC AAB36354;  
 DT 26-FEB-2001 (first entry)  
 XX Agfa::PT3#9 amino acid sequence SEQ ID NO:28.  
 DE Salmorella; agfa; chromosomal gene replacement; fimbria; epitope;  
 KW vaccine; immune response; immunogen.  
 XX Salmorella enteritidis.  
 OS Escherichia coli.  
 OS Synthetic.  
 XX WO200060102-A2.  
 PN 12-OCT-2000.  
 XX 05-APR-2000; 2000WO-CA000356.  
 XX 05-APR-1999; 99US-0127888P.  
 XX (UYVI-) UNIV VICTORIA.  
 XX White AP, Doran JL, Collison SK, Kay WW;  
 PI WPI; 2000-672631/65.  
 DR N-PSDB; AAC64630.  
 XX Recombinant agfa gene having a segment replaced by a foreign DNA sequence  
 PT which encodes foreign epitope or antigen, expresses recombinant Agfa  
 PT protein useful for eliciting immune response in animal.  
 XX Disclosure; Page 138; 139pp; English.

XX The present invention describes a recombinant agfa gene (I) where a  
 CC segment of the gene has been replaced by a segment of a foreign DNA  
 CC sequence which encodes a foreign epitope or antigen. Also described are:  
 CC (1) use of thin aggregative fimbriae (SEF17/TAF) nucleation depended  
 CC assembly system of strains of Salmonella, Escherichia coli and  
 CC Enterobacteriaceae for the production of fimbriae comprising recombinant  
 CC Agfa, CsgA and Agfa-homologue fimbrial subunits, respectively; (2)  
 CC directing recombination of a recombinant gene into the chromosome of the  
 CC homologous species; (3) directing recombination of a recombinant gene  
 CC back into the chromosome of the homologous species, replacing the native  
 CC copy of that gene; and (4) eliciting an immune response in an animal,  
 CC comprising separating an amino acid polymer comprising a recombinant Agfa  
 CC protein containing a replacement segment or segments of foreign amino  
 CC acid sequence or sequences grown on a Salmonella, E. coli or  
 CC Enterobacteriaceae host cell, from the host cell and introducing the  
 CC polymer into the animal in conjunction with a carrier or diluent. (I) is  
 CC useful for the expression of recombinant Agfa protein which is useful for  
 CC eliciting an immune response in an animal. In a fimbrial presentation  
 CC system the heterologous antigens are presented in high numbers (up to  
 CC 500,000 copies/cell), the hybrid fimbrial protein possesses both the  
 CC immunogenicity and adhesion properties relevant for an efficient live  
 CC vaccine, the carrier fimbrial subunit proteins are usually strong  
 CC immunogens, which may be important for directing an immune response  
 CC against the inserted epitope, and hybrid fimbriae are easy and  
 CC inexpensive to purify in large amount. The present sequence is given in  
 CC the exemplification of the present invention  
 XX  
 SQ Sequence 151 AA;

Query Match 91.8%; Score 712; DB 3; Length 151;  
 Best Local Similarity 91.1%; Pred. No. 1.7e-61;  
 Matches 144; Conservative 0; Mismatches 0; Indels 14; Gaps 2;  
 QY 1 MLLKLVAAFAAIVVSGSALAGVVPQWGGGNGHNGGNSGPDSTLSIYQYGSANAALALQ 60  
 DB 1 MLLKLVAAFAAIVVSGSALAGVVPQWGGGNGHNGGNSGPDSTLSIYQYGSANAALALQ 60  
 QY 61 SDARKSETTITQSGYNGADVGQADNYDOLVTRVVTHEMAHA-----DOWNAKNSDI 113  
 DB 61 SDARKSETTITQSGYNGAD-----YDOLVTRVVTHEMAHAFRNNATIDOWNAKNSDI 113  
 QY 114 TVGQYGGNNAALVNTASDSSVMVRQVGFNNATANQY 151  
 DB 114 TVGQYGGNNAALVNTASDSSVMVRQVGFNNATANQY 151

RESULT 3  
 AAR74625  
 ID AAR74625 standard; protein; 151 AA.  
 AC AAR74625;  
 DT 25-MAR-2003 (revised)  
 DT 26-JUN-1995 (first entry)  
 XX Agfa sequence.  
 DE Salmorella; Agfa; vaccine.  
 KW Salmorella.  
 OS Salmorella.  
 PN WO9425598-A2.  
 XX 10-NOV-1994.  
 XX 26-APR-1994; 94WO-IB000207.  
 XX 26-APR-1993; 93US-00054452.  
 XX (UYVI-) UNIV VICTORIA INNOVATION & DEV CORP.  
 PA (KING/) KING J.  
 XX

PI Kay WW, Collinson SK, Clouthier SC, Doran JL;  
XX WPI; 1994-358275/44.  
DR N-PSDB; AAG87467.  
XX  
PT Eliciting an immune response to Salmonella - using attenuated Salmonella  
PT strains, vector constructs, or compsns. contg. fimbrial type proteins.  
XX  
PS Disclosure; Fig 7B; 95pp; English.  
XX  
XX The Salmonella Agfa protein and DNA are used in vaccine and genetic  
CC immunization compositions, respectively, to elicit an immune response to  
CC Salmonella in animals (e.g. food producing animals) and humans. (Updated  
CC on 25-MAR-2003 to correct PN field.)  
XX  
SQ Sequence 151 AA;  
Query Match 87.9%; Score 682; DB 2; Length 151;  
Best Local Similarity 90.7%; Pred. No. 1.4e-58;  
Matches 137; Conservative 2; Mismatches 12; Indels 0; Gaps 0;  
OY 1 MKLLKVAFAAIVVSGSALAGVVPQWGGGNGGNSGPGDSTLSIYQYGSANAALALQ 60  
DB 1 MKLLKVAFAAIVVSGSALAGVVPQWGGGNGGNSGPGDSTLSIYQYGSANAALALQ 60  
OY 61 SDARKSETTITQSGYNGADVGQADNYDQLVTRVTHEMAHADQWNAKNSDITVGYGG 120  
DB 61 SDARKSETTITQSGYNGADVGQADNSTIETLQNGFRNNATIDQWNAKNSDITVGYGG 120  
OY 121 NNAALVNQATSDSSVMVRQVGFNNATANQY 151  
DB 121 NNAALVNQATSDSSVMVRQVGFNNATANQY 151  
RESULT 4  
AAB36341  
ID AAB36341 standard; protein; 151 AA.  
AC AAB36341;  
XX  
XX 26-FEB-2001 (first entry)  
XX Salmonella enteritidis Agfa amino acid sequence SEQ ID NO:5.  
DE DE  
XX Salmonella; agfa; chromosomal gene replacement; fimbria; epitope;  
KW vaccine; immune response; immunogen.  
XX  
XX Salmonella enteritidis.  
XX  
XX W0200060102-A2.  
XX  
XX 12-OCT-2000.  
XX  
XX 05-APR-2000; 2000WO-CA000356.  
XX  
XX 05-APR-1999; 99US-0127888P.  
XX  
XX (UUVI-) UNIV VICTORIA.  
XX  
XX White AP, Doran JL, Collinson SK, Kay WW;  
PI  
XX WPI; 2000-672631/65.  
DR  
DR N-PSDB; AAC64617.  
XX  
XX Recombinant agfa gene having a segment replaced by a foreign DNA sequence  
PT which encodes foreign epitope or antigen, expresses recombinant Agfa  
PT protein useful for eliciting immune response in animal.  
XX  
XX Disclosure; Page 135; 139pp; English.  
XX  
XX The present invention describes a recombinant agfa gene (I) where a  
CC segment of the gene has been replaced by a segment of a foreign DNA  
CC sequence which encodes a foreign epitope or antigen. Also described are:

CC (1) use of thin aggregative fimbriae (SEF17/TAF) nucleation depended  
CC assembly system of strains of Salmonella, Escherichia coli and  
CC Enterobacteriaceae for the production of fimbriae comprising recombinant  
CC Agfa, CsgA and Agfa-homologue fimbria subunits, respectively; (2)  
CC directing recombination of a recombinant gene into the chromosome of the  
CC homologous species; (3) directing recombination of a recombinant gene  
CC back into the chromosome of the homologous species, replacing the native  
CC copy of that gene; and (4) eliciting an immune response in an animal,  
CC comprising separating an amino acid polymer comprising a recombinant Agfa  
CC protein containing a replacement segment or segments of foreign amino  
CC acid sequence or sequences grown on a Salmonella, E. coli or  
CC Enterobacteriaceae host cell, from the host cell and introducing the  
CC polymer into the animal in conjunction with a carrier or diluent. (I) is  
CC useful for the expression of recombinant Agfa protein which is useful for  
CC eliciting an immune response in an animal. In a fimbrial presentation  
CC system the heterologous antigens are presented in high numbers (up to  
CC 500,000 copies/cell), the hybrid fimbria protein possesses both the  
CC immunogenicity and adhesion properties relevant for an efficient live  
CC vaccine, the carrier fimbrial subunit proteins are usually strong  
CC immunogens, which may be important for directing an immune response  
CC against the inserted epitope, and hybrid fimbriae are easy and  
CC inexpensive to purify in large amount. The present sequence is given in  
CC the exemplification of the present invention  
XX  
SQ Sequence 151 AA;  
Query Match 87.9%; Score 682; DB 3; Length 151;  
Best Local Similarity 90.7%; Pred. No. 1.4e-58;  
Matches 137; Conservative 2; Mismatches 12; Indels 0; Gaps 0;  
OY 1 MKLLKVAFAAIVVSGSALAGVVPQWGGGNGGNSGPGDSTLSIYQYGSANAALALQ 60  
DB 1 MKLLKVAFAAIVVSGSALAGVVPQWGGGNGGNSGPGDSTLSIYQYGSANAALALQ 60  
OY 61 SDARKSETTITQSGYNGADVGQADNYDQLVTRVTHEMAHADQWNAKNSDITVGYGG 120  
DB 61 SDARKSETTITQSGYNGADVGQADNSTIETLQNGFRNNATIDQWNAKNSDITVGYGG 120  
OY 121 NNAALVNQATSDSSVMVRQVGFNNATANQY 151  
DB 121 NNAALVNQATSDSSVMVRQVGFNNATANQY 151  
RESULT 5  
AAW23570  
ID AAW23570 standard; protein; 151 AA.  
XX  
XX AAW23570;  
XX  
XX 25-MAR-2003 (revised)  
DT 29-SEP-1997 (first entry)  
XX  
XX Salmonella enteritidis 27655-3b agfa.  
XX Enteropathogenic bacteria; enterobacteria; S. enteritidis; antibody.  
XX Salmonella enteritidis.  
XX  
XX Key Location/Qualifiers  
FH Misc-difference 123  
FT /note= "Encoded by GCC"  
XX  
XX US5635617-A.  
XX  
XX 03-JUN-1997.  
XX  
XX 26-APR-1994; 94US-00233788.  
XX  
XX 26-APR-1993; 93US-00054452.  
XX  
XX (UUVI-) UNIV VICTORIA INNOVATION & DEV CORP.  
XX Collinson SK, Kay WW, Doran JL;  
PI

XX WPI: 1997-309886/28.  
 DR N-PSDB; AAT74142.  
 XX  
 XX Isolated Salmonella gene agfa - used for diagnosis of Salmonella or  
 PT enteropathogenic bacteria of the Enterobacteria family.  
 XX  
 XX Example 2; Fig 7; 85pp; English.  
 XX  
 CC The present sequence represents agfa encoded by the full agfa gene  
 CC derived from Salmonella enteritidis 27855-3b. The nucleic acid can be  
 CC used to provide diagnostic assays for Salmonella and/or enteropathogenic  
 CC bacteria of the family Enterobacteria. It can also be used to provide  
 CC proteins and antibodies which can be used for assays. The nucleic acid  
 CC sequence can be used to provide probes or primers which can specifically  
 CC hybridise to nucleic acid molecules from greater than 99% of Salmonella  
 CC strains that are pathogenic to warm-blooded animals relative to nucleic  
 CC acid molecules from virtually all other microbial organisms. (Updated on  
 CC 25-MAR-2003 to correct PF field.)  
 XX  
 SQ Sequence 151 AA;  
 Query Match 87.2%; Score 677; DB 2; Length 151;  
 Best Local Similarity 90.1%; Pred. No. 4.4e-58;  
 Matches 136; Conservative 2; Mismatches 13; Indels 0; Gaps 0;  
 QY 1 MKLLKVAAPAAIYVSGSALAGVVPQGGGNGHNGGNSGPDSTLSIYQYGSANAALALQ 60  
 DB 1 MKLLKVAAPAAIYVSGSALAGVVPQGGGNGHNGGNSGPDSTLSIYQYGSANAALALQ 60  
 QY 61 SDARKSETTITQSGYNGADVGQADNYDQVTRVVTHEMAHADQWAKNSDITVGOYGG 120  
 DB 61 SDARKSETTITQSGYNGADVGQADNSTIELTQNGFRNNATIDQWAKNSDITVGOYGG 120  
 QY 121 NNAALVNQTASDSSVMVRQVFGNNATANQY 151  
 DB 121 NNPAALVNQTASDSSVMVRQVFGNNATANQY 151  
 RESULT 6  
 AAB36355  
 ID AAB36355 standard; protein; 151 AA.  
 AC AAB36355;  
 XX  
 DT 26-FEB-2001 (first entry)  
 DE Agfa::PT3#10 amino acid sequence SEQ ID NO:30.  
 XX  
 XX Salmonella; agfa; chromosomal gene replacement; fimbrin; epitope;  
 KW vaccine; immune response; immunogen.  
 XX  
 OS Salmonella enteritidis.  
 OS Escherichia coli.  
 OS Synthetic.  
 XX  
 PN WO200060102-A2.  
 XX  
 PD 12-OCT-2000.  
 XX  
 PF 05-APR-2000; 2000WO-CA000356.  
 XX  
 PR 05-APR-1999; 99US-0127888P.  
 XX  
 PA (UYVI-) UNIV VICTORIA.  
 XX  
 PI White AP, Doran JL, Collison SK, Kay WW;  
 XX  
 XX WPI: 2000-672631/65.  
 DR N-PSDB; AAC64631.  
 XX  
 PT Recombinant agfa gene having a segment replaced by a foreign DNA sequence  
 PT which encodes foreign epitope or antigen, expresses recombinant Agfa

PT protein useful for eliciting immune response in animal.  
 XX  
 PS Disclosure; Page 139; 139pp; English.  
 XX  
 CC The present invention describes a recombinant agfa gene (I) where a  
 CC segment of the gene has been replaced by a segment of a foreign DNA  
 CC sequence which encodes a foreign epitope or antigen. Also described are:  
 CC (1) use of thin aggregative fimbriae (SEF17/7AF) nucleation depended are:  
 CC assembly system of strains of Salmonella, Escherichia coli and  
 CC Enterobacteriaceae for the production of fimbriae comprising recombinant  
 CC Agfa, CsgA and Agfa-homologue fimbrial subunits, respectively; (2)  
 CC directing recombination of a recombinant gene into the chromosome of the  
 CC homologous species; (3) directing recombination of a recombinant gene  
 CC back into the chromosome of the homologous species, replacing the native  
 CC copy of that gene; and (4) eliciting an immune response in an animal,  
 CC comprising separating an amino acid polymer comprising a recombinant Agfa  
 CC protein containing a replacement segment or segments of foreign amino  
 CC acid sequence or sequences grown on a Salmonella, E. coli or  
 CC Enterobacteriaceae host cell, from the host cell and introducing the  
 CC polymer into the animal in conjunction with a carrier or diluent. (I) is  
 CC useful for the expression of recombinant Agfa protein which is useful for  
 CC eliciting an immune response in an animal. In a fimbrial presentation  
 CC system the heterologous antigens are presented in high numbers (up to  
 CC 500,000 copies/cell), the hybrid fimbrial protein possesses both the  
 CC immunogenicity and adhesion properties relevant for an efficient live  
 CC vaccine, the carrier fimbrial subunit proteins are usually strong  
 CC immunogens, which may be important for directing an immune response  
 CC against the inserted epitope, and hybrid fimbriae are easy and  
 CC inexpensive to purify in large amount. The present sequence is given in  
 CC the exemplification of the present invention  
 XX  
 SQ Sequence 151 AA;  
 Query Match 84.7%; Score 657; DB 3; Length 151;  
 Best Local Similarity 81.9%; Pred. No. 4e-56;  
 Matches 136; Conservative 0; Mismatches 0; Indels 30; Gaps 2;  
 QY 1 MKLLKVAAPAAIYVSGSALAGVVPQGGGNGHNGGNSGPDSTLSIYQYGSANAALALQ 60  
 DB 1 MKLLKVAAPAAIYVSGSALAGVVPQGGGNGHNGGNSGPDSTLSIYQYGSANAALALQ 60  
 QY 61 SDARKSETTITQSGYNGADVGQADN-----YDQLVTRVVTHEMAHADQ 105  
 DB 61 SDARKSETTITQSGYNGADVGQADNSTIELTQNGFRNNATYDQLVTRVVTHEMAHA-- 118  
 QY 106 WNAKNSDITVGOYGGNNAALVNQTASDSSVMVRQVFGNNATANQY 151  
 DB 119 -----GGNNAALVNQTASDSSVMVRQVFGNNATANQY 151  
 RESULT 7  
 AAB36350  
 ID AAB36350 standard; protein; 151 AA.  
 AC AAB36350;  
 XX  
 DT 26-FEB-2001 (first entry)  
 DE Agfa::PT3#5 amino acid sequence SEQ ID NO:20.  
 XX  
 KW Salmonella; agfa; chromosomal gene replacement; fimbrin; epitope;  
 KW vaccine; immune response; immunogen.  
 XX  
 OS Salmonella enteritidis.  
 OS Escherichia coli.  
 OS Synthetic.  
 XX  
 PN WO200060102-A2.  
 XX  
 PD 12-OCT-2000.  
 XX  
 PF 05-APR-2000; 2000WO-CA000356.  
 XX

PR 05-APR-1999; 99US-0127888P.  
XX (UYVI-) UNIV VICTORIA.  
XX White AP, Doran JL, Collison SK, Kay WW;  
XX WPI: 2000-672631/65.  
DR N-PSDB; AAC64626.  
XX Recombinant agfA gene having a segment replaced by a foreign DNA sequence  
PT which encodes foreign epitope or antigen, expresses recombinant AgfA  
PT protein useful for eliciting immune response in animal.  
XX Disclosure; Page 137; 139pp; English.  
XX The present invention describes a recombinant agfA gene (I) where a  
CC segment of the gene has been replaced by a segment of a foreign DNA  
CC sequence which encodes a foreign epitope or antigen. Also described are:  
CC (1) use of thin aggregative fimbriae (SEF17/TAF) nucleation depended  
CC assembly system of strains of Salmonella, Escherichia coli and  
CC Enterobacteriaceae for the production of fimbriae comprising recombinant  
CC AgfA, CsgA and AgfA-homologue fimbrial subunits, respectively; (2)  
CC directing recombination of a recombinant gene into the chromosome of the  
CC homologous species; (3) directing recombination of a recombinant gene  
CC back into the chromosome of the homologous species, replacing the native  
CC copy of that gene; and (4) eliciting an immune response in an animal,  
CC comprising separating an amino acid polymer comprising a recombinant AgfA  
CC protein containing a replacement segment or segments of foreign amino  
CC acid sequence or sequences grown on a Salmonella, E. coli or  
CC Enterobacteriaceae host cell, from the host cell and introducing the  
CC polymer into the animal in conjunction with a carrier or diluent. (I) is  
CC useful for the expression of recombinant AgfA protein which is useful for  
CC eliciting an immune response in an animal. In a fimbrial presentation  
CC system the heterologous antigens are presented in high numbers (up to  
CC 500,000 copies/cell), the hybrid fimbrial protein possesses both the  
CC immunogenicity and adhesion properties relevant for an efficient live  
CC vaccine, the carrier fimbrial subunit proteins are usually strong  
CC immunogens, which may be important for directing an immune response  
CC against the inserted epitope, and hybrid fimbriae are easy and  
CC inexpensive to purify in large amount. The present sequence is given in  
CC the exemplification of the present invention  
XX Sequence 151 AA;  
SQ  
Query Match 79.8%; Score 619; DB 3; Length 151;  
Best Local Similarity 74.6%; Pred. No. 2e-52;  
Matches 129; Conservative 0; Mismatches 0; Indels 44; Gaps 2;  
QY 1 MKLLKVAAPAAIVVSGSALAGVVPQGGGNGHNGGSGDPDSTLSIYQYGSANAALALQ 60  
DB 1 MKLLKVAAPAAIVVSGSALAGVVPQGGGNGHNGGSGDPDSTLSIYQYGSANAALALQ 60  
QY 61 SDARKSETTITQSGYNGADVGQADNYDQLVTRVTVTHEMAHA-----103  
DB 61 SDARK-----YDQLVTRVTVTHEMAHAGQADNSTIELTQNGFR 98  
QY 104 -----DQNAKNKSDITVQYGGNNAALVNQTSADSSVMVRQVGFNNATANQY 151  
DB 99 NNATIDQNAKNKSDITVQYGGNNAALVNQTSADSSVMVRQVGFNNATANQY 151  
RESULT 8  
AAB36352  
XX AAB36352 standard; protein; 151 AA.  
XX AC AAB36352;  
XX 26-FEB-2001 (first entry)  
XX AgfA::PT3#7 amino acid sequence SEQ ID NO:24.  
XX DE Salmonella; agfA; chromosomal gene replacement; fimbrial; epitope;  
XX KW vaccine; immune response; immunogen.

XX Salmonella enteritidis.  
OS Escherichia coli.  
OS Synthetic.  
XX WO2000060102-A2.  
XX 12-OCT-2000.  
XX 05-APR-2000; 2000WO-CA000356.  
XX 05-APR-1999; 99US-0127888P.  
XX (UYVI-) UNIV VICTORIA.  
XX White AP, Doran JL, Collison SK, Kay WW;  
XX WPI: 2000-672631/65.  
DR N-PSDB; AAC64626.  
XX Recombinant agfA gene having a segment replaced by a foreign DNA sequence  
PT which encodes foreign epitope or antigen, expresses recombinant AgfA  
PT protein useful for eliciting immune response in animal.  
XX Disclosure; Page 138; 139pp; English.  
XX The present invention describes a recombinant agfA gene (I) where a  
CC segment of the gene has been replaced by a segment of a foreign DNA  
CC sequence which encodes a foreign epitope or antigen. Also described are:  
CC (1) use of thin aggregative fimbriae (SEF17/TAF) nucleation depended  
CC assembly system of strains of Salmonella, Escherichia coli and  
CC Enterobacteriaceae for the production of fimbriae comprising recombinant  
CC AgfA, CsgA and AgfA-homologue fimbrial subunits, respectively; (2)  
CC directing recombination of a recombinant gene into the chromosome of the  
CC homologous species; (3) directing recombination of a recombinant gene  
CC back into the chromosome of the homologous species, replacing the native  
CC copy of that gene; and (4) eliciting an immune response in an animal,  
CC comprising separating an amino acid polymer comprising a recombinant AgfA  
CC protein containing a replacement segment or segments of foreign amino  
CC acid sequence or sequences grown on a Salmonella, E. coli or  
CC Enterobacteriaceae host cell, from the host cell and introducing the  
CC polymer into the animal in conjunction with a carrier or diluent. (I) is  
CC useful for the expression of recombinant AgfA protein which is useful for  
CC eliciting an immune response in an animal. In a fimbrial presentation  
CC system the heterologous antigens are presented in high numbers (up to  
CC 500,000 copies/cell), the hybrid fimbrial protein possesses both the  
CC immunogenicity and adhesion properties relevant for an efficient live  
CC vaccine, the carrier fimbrial subunit proteins are usually strong  
CC immunogens, which may be important for directing an immune response  
CC against the inserted epitope, and hybrid fimbriae are easy and  
CC inexpensive to purify in large amount. The present sequence is given in  
CC the exemplification of the present invention  
XX Sequence 151 AA;  
SQ  
Query Match 78.5%; Score 609; DB 3; Length 151;  
Best Local Similarity 73.6%; Pred. No. 1.9e-51;  
Matches 128; Conservative 0; Mismatches 0; Indels 46; Gaps 2;  
QY 1 MKLLKVAAPAAIVVSGSALAGVVPQGGGNGHNGGSGDPDSTLSIYQYGSANAALALQ 60  
DB 1 MKLLKVAAPAAIVVSGSALAGVVPQGGGNGHNGGSGDPDSTLSIYQYGSANAALALQ 60  
QY 61 SDARKSETTITQSGYNGADVGQAD-----NYDQLVTRVTV 97  
DB 61 SDARKSETTITQSGYNGADVGQADNSTIELTQNGFRNATIDQNAKNYDQLVTRVTV 120  
QY 98 HEMAHADQNAKNKSDITVQYGGNNAALVNQTSADSSVMVRQVGFNNATANQY 151  
DB 121 HEMAH-----NQTSADSSVMVRQVGFNNATANQY 151  
RESULT 9

AAB36346  
 ID AAB36346 standard; protein; 151 AA.  
 AC AAB36346;  
 XX  
 XX 26-FEB-2001 (first entry)  
 DT  
 DT  
 DE  
 DE Agfa::PT3#1 amino acid sequence SEQ ID NO:12.  
 XX  
 XX Salmonella; agfa; chromosomal gene replacement; fimbrin; epitope;  
 KW vaccine; immune response; immunogen.  
 XX  
 XX Salmonella enteritidis.  
 OS Escherichia coli.  
 OS Synthetic.  
 XX  
 PN WO200060102-A2.  
 XX  
 PD 12-OCT-2000.  
 XX  
 XX 05-APR-2000; 2000WO-CA000356.  
 XX  
 XX 05-APR-1999; 99US-0127888P.  
 PR  
 XX (UYVI-) UNIV VICTORIA.  
 PA  
 PA White AP, Doran JL, Collison SK, Kay WW;  
 XX  
 XX WPI; 2000-672631/65.  
 DR N-PSDB; AAC64622.  
 DR  
 XX  
 PT Recombinant agfa gene having a segment replaced by a foreign DNA sequence  
 PT which encodes foreign epitope or antigen, expresses recombinant Agfa  
 PT protein useful for eliciting immune response in animal.  
 XX  
 XX Disclosure; Page 135; 139pp; English.  
 XX  
 CC The present invention describes a recombinant agfa gene (I) where a  
 CC segment of the gene has been replaced by a segment of a foreign DNA  
 CC sequence which encodes a foreign epitope or antigen. Also described are:  
 CC (1) use of thin aggregative fimbriae (SEF17/TAF) nucleation depended  
 CC assembly system of strains of Salmonella, Escherichia coli and  
 CC Enterobacteriaceae for the production of fimbriae comprising recombinant  
 CC Agfa, CsgA and Agfa-homologue fimbrin subunits, respectively; (2)  
 CC directing recombination of a recombinant gene into the chromosome of the  
 CC homologous species; (3) directing recombination of a recombinant gene  
 CC back into the chromosome of the homologous species, replacing the native  
 CC copy of that gene; and (4) eliciting an immune response in an animal,  
 CC comprising separating an amino acid polymer comprising a recombinant Agfa  
 CC protein containing a replacement segment or segments of foreign amino  
 CC acid sequence or sequences grown on a Salmonella, E. coli or  
 CC Enterobacteriaceae host cell, from the host cell and introducing the  
 CC polymer into the animal in conjunction with a carrier or diluent. (I) is  
 CC useful for the expression of recombinant Agfa protein which is useful for  
 CC eliciting an immune response in an animal. In a fimbrial presentation  
 CC system the heterologous antigens are presented in high numbers (up to  
 CC 500,000 copies/cell) the hybrid fimbrin protein possesses both the  
 CC immunogenicity and adhesion properties relevant for an efficient live  
 CC vaccine, the carrier fimbrial subunit proteins are usually strong  
 CC immunogens, which may be important for directing an immune response  
 CC against the inserted epitope, and hybrid fimbriae are easy and  
 CC inexpensive to purify in large amount. The present sequence is given in  
 CC the exemplification of the present invention  
 XX  
 XX Sequence 151 AA;  
 SQ  
 Query Match 77.8%; Score 604; DB 3; Length 151;  
 Best Local Similarity 80.8%; Pred. No. 5.9e-51;  
 Matches 122; Conservative 6; Mismatches 23; Indels 0; Gaps 0;  
 1 MKLLKVAFAAIVVSGSALAGVVPQGGGNGGNGSGPDSSTLSIYQGSANALALQ 60  
 1 MKLLKVAFAAIVVSGSALAGVVPQGGGNGGNGSGPDSSTLSIYQGSANALALQ 60

CC the exemplification of the present invention  
 XX Sequence 151 AA;  
 SQ

Query Match 77.6%; Score 602; DB 3; Length 151;  
 Best Local Similarity 81.5%; Pred. No. 9.3e-51;  
 Matches 123; Conservative 23; Mismatches 23; Indels 0; Gaps 0;

QY 1 MLLKVAFAAIVVSGSALAGVVPQWGGGNGGNSGPDSTLSIYQGSANAALALQ 60  
 DB 1 MLLKVAFAAIVVSGSALAGVVPQWGGGNGGNSGPDSTLSIYQGSANAALALQ 60  
 QY 61 SPARKSETTITQSGYNGADVGQADNYDQLVTRVVTHEMAHADQWNAKNSDITVGQYGG 120  
 DB 61 SPARKSETTITQSGYNGADVGQADNSTIELTQNGFRNNATIDQWNAKNSDITVGQYDQ 120  
 QY 121 NNAALVNQATSDSSVMVRQVGFNNATANQY 151  
 DB 121 LVTRVVTHEMAHASVMVRQVGFNNATANQY 151

RESULT 11  
 AAB36353  
 ID AAB36353 standard; protein; 151 AA.  
 XX AC AAB36353;  
 XX DT 26-FEB-2001 (first entry)  
 XX DE Agfa::PT3#8 amino acid sequence SEQ ID NO:26.  
 XX KW Salmonella; agfa; chromosomal gene replacement; fimbria; epitope;  
 XX KW vaccine; immune response; immunogen.  
 XX OS Salmonella enteritidis.  
 XX OS Escherichia coli.  
 XX OS Synthetic.  
 XX PN W0200060102-A2.  
 XX PD 12-OCT-2000.  
 XX PF 05-APR-2000; 2000WO-CA000356.  
 XX PR 05-APR-1999; 99US-0127888P.  
 XX PA (UYVI-) UNIV VICTORIA.  
 XX PI White AP, Doran JL, Collison SK, Kay WW;  
 XX DR WPI: 2000-672631/65.  
 XX DR N-PSDB; AAC64625.  
 XX PT Recombinant agfa gene having a segment replaced by a foreign DNA sequence  
 XX PT which encodes foreign epitope or antigen, expresses recombinant Agfa  
 XX PT protein useful for eliciting immune response in animal.  
 XX PS Disclosure; Page 138; 139pp; English.  
 XX PI The present invention describes a recombinant agfa gene (I) where a  
 XX CC segment of the gene has been replaced by a segment of a foreign DNA  
 XX CC sequence which encodes a foreign epitope or antigen. Also described are:  
 XX CC (1) use of thin aggregative fimbriae (SEF17/TAF) nucleation depended  
 XX CC assembly system of strains of Salmonella, Escherichia coli and  
 XX CC Enterobacteriaceae for the production of fimbriae comprising recombinant  
 XX CC Agfa, CsgA and Agfa-homologue fimbria subunits, respectively; (2)  
 XX CC directing recombination of a recombinant gene into the chromosome of the  
 XX CC homologous species; (3) directing recombination of a recombinant gene  
 XX CC back into the chromosome of the homologous species, replacing the native  
 XX CC copy of that gene; and (4) eliciting an immune response in an animal,  
 XX CC comprising separating an amino acid polymer comprising a recombinant Agfa  
 XX CC protein containing a replacement segment or segments of foreign amino  
 XX CC acid sequence or sequences grown on a Salmonella, E. coli or

CC Enterobacteriaceae host cell, from the host cell and introducing the  
 CC polymer into the animal in conjunction with a carrier or diluent. (I) is  
 CC useful for the expression of recombinant Agfa protein which is useful for  
 CC eliciting an immune response in an animal. In a fimbrial presentation  
 CC system the heterologous antigens are presented in high numbers (up to  
 CC 500,000 copies/cell), the hybrid fimbria protein possesses both the  
 CC immunogenicity and adhesion properties relevant for an efficient live  
 CC vaccine, the carrier fimbrial subunit proteins are usually strong  
 CC immunogens, which may be important for directing an immune response  
 CC against the inserted epitope, and hybrid fimbriae are easy and  
 CC inexpensive to purify in large amount. The present sequence is given in  
 CC the exemplification of the present invention  
 XX Sequence 151 AA;  
 SQ

Query Match 77.4%; Score 601; DB 3; Length 151;  
 Best Local Similarity 81.5%; Pred. No. 1.2e-50;  
 Matches 123; Conservative 4; Mismatches 24; Indels 0; Gaps 0;

QY 1 MLLKVAFAAIVVSGSALAGVVPQWGGGNGGNSGPDSTLSIYQGSANAALALQ 60  
 DB 1 MLLKVAFAAIVVSGSALAGVVPQWGGGNGGNSGPDSTLSIYQGSANAALALQ 60  
 QY 61 SPARKSETTITQSGYNGADVGQADNYDQLVTRVVTHEMAHADQWNAKNSDITVGQYGG 120  
 DB 61 LVTRVVTHEMAHAGYNGADVGQADNSTIELTQNGFRNNATIDQWNAKNSDITVGQYGG 120  
 QY 121 NNAALVNQATSDSSVMVRQVGFNNATANQY 151  
 DB 121 NNAALVNQATSDSSVMVRQVGFNNATANQY 151

RESULT 12  
 AAB36349  
 ID AAB36349 standard; protein; 151 AA.  
 XX AC AAB36349;  
 XX DT 26-FEB-2001 (first entry)  
 XX DE Agfa::PT3#4 amino acid sequence SEQ ID NO:18.  
 XX KW Salmonella; agfa; chromosomal gene replacement; fimbria; epitope;  
 XX KW vaccine; immune response; immunogen.  
 XX OS Salmonella enteritidis.  
 XX OS Escherichia coli.  
 XX OS Synthetic.  
 XX PN W0200060102-A2.  
 XX PD 12-OCT-2000.  
 XX PF 05-APR-2000; 2000WO-CA000356.  
 XX PR 05-APR-1999; 99US-0127888P.  
 XX PA (UYVI-) UNIV VICTORIA.  
 XX PI White AP, Doran JL, Collison SK, Kay WW;  
 XX DR WPI: 2000-672631/65.  
 XX DR N-PSDB; AAC64625.  
 XX PT Recombinant agfa gene having a segment replaced by a foreign DNA sequence  
 XX PT which encodes foreign epitope or antigen, expresses recombinant Agfa  
 XX PT protein useful for eliciting immune response in animal.  
 XX PS Disclosure; Page 136; 139pp; English.  
 XX PI The present invention describes a recombinant agfa gene (I) where a  
 XX CC segment of the gene has been replaced by a segment of a foreign DNA  
 XX CC sequence which encodes a foreign epitope or antigen. Also described are:

CC (1) use of thin aggregative fimbriae (SEF17/TAF) nucleation depended  
 CC assembly system of strains of Salmonella, Escherichia coli and  
 CC Enterobacteriaceae for the production of fimbriae comprising recombinant  
 CC Agfa, CsgA and Agfa-homologue fimbrial subunits, respectively; (2)  
 CC directing recombination of a recombinant gene into the chromosome of the  
 CC homologous species; (3) directing recombination of a recombinant gene  
 CC back into the chromosome of the homologous species, replacing the native  
 CC copy of that gene; and (4) eliciting an immune response in an animal,  
 CC comprising separating an amino acid polymer comprising a recombinant Agfa  
 CC protein containing a replacement segment or segments of foreign amino  
 CC acid sequence or sequences grown on a Salmonella, E. coli or  
 CC Enterobacteriaceae host cell, from the host cell and introducing the  
 CC polymer into the animal in conjunction with a carrier or diluent. (1) is  
 CC useful for the expression of recombinant Agfa protein which is useful for  
 CC eliciting an immune response in an animal. In a fimbrial presentation  
 CC system the heterologous antigens are presented in high numbers (up to  
 CC 500,000 copies/cell), the hybrid fimbrial protein possesses both the  
 CC immunogenicity and adhesion properties relevant for an efficient live  
 CC vaccine, the carrier fimbrial subunit proteins are usually strong  
 CC immunogens, which may be important for directing an immune response  
 CC against the inserted epitope, and hybrid fimbriae are easy and  
 CC inexpensive to purify in large amount. The present sequence is given in  
 CC the exemplification of the present invention

XX Sequence 151 AA;

Query Match 77.3%; Score 600; DB 3; Length 151;  
 Best Local Similarity 81.5%; Pred. No. 1.5e-50;  
 Matches 123; Conservative 5; Mismatches 23; Indels 0; Gaps 0;  
 QY 1 MLLKVAAPAAIVVSGSALAGVVPQGGGNGHNGSGPDSLTSTIYQYGSANAALALQ 60  
 DB 1 MLLKVAAPAAIVVSGSALAGVVPQGGGNGHNGSGPDSLTSTIYQYGSANAALALQ 60  
 QY 61 SDARKSETTTTQSGYNGADVGADNYDQLVTRVVTHEMAHADQNAKNSDITVQYGG 120  
 DB 61 SDARKSETTTTQSGYNGADVGADNYDQLVTRVVTHEMAHADQNAKNSDITVQYGG 120  
 QY 121 NNAALVNQTASDSSVMVRQVFGNNATANQY 151  
 DB 121 NNAALVNQTASDSSVMVRQVFGNNATANQY 151

RESULT 13

AAB36348  
 ID AAB36348 standard; protein; 151 AA.

XX AAB36348;

XX 26-FEB-2001 (first entry)

DE Agfa::PT3#3 amino acid sequence SEQ ID NO:16.

XX Salmonella; agfa; chromosomal gene replacement; fimbrial; epitope;  
 KW vaccine; immune response; immunogen.

XX Salmonella enteritidis.

OS Escherichia coli.

OS Synthetic.

XX WO2000060102-A2.

PN 12-OCT-2000.

XX 05-APR-2000; 2000WO-CA000356.

XX 05-APR-1999; 99US-0127888P.

XX (UyVI-) UNIV VICTORIA.

XX White AP, Doran JL, Collison SK, Kay WW;

XX WPI; 2000-672631/65.

DR N-PSDB; AAC64624.

XX Recombinant agfa gene having a segment replaced by a foreign DNA sequence  
 PT which encodes foreign epitope or antigen, expresses recombinant Agfa  
 PT protein useful for eliciting immune response in animal.

XX Disclosure; Page 136; 139pp; English.

XX The present invention describes a recombinant agfa gene (I) where a  
 CC segment of the gene has been replaced by a segment of a foreign DNA  
 CC sequence which encodes a foreign epitope or antigen. Also described are:  
 CC (1) use of thin aggregative fimbriae (SEF17/TAF) nucleation depended  
 CC assembly system of strains of Salmonella, Escherichia coli and  
 CC Enterobacteriaceae for the production of fimbriae comprising recombinant  
 CC Agfa, CsgA and Agfa-homologue fimbrial subunits, respectively; (2)  
 CC directing recombination of a recombinant gene into the chromosome of the  
 CC homologous species; (3) directing recombination of a recombinant gene  
 CC back into the chromosome of the homologous species, replacing the native  
 CC copy of that gene; and (4) eliciting an immune response in an animal,  
 CC comprising separating an amino acid polymer comprising a recombinant Agfa  
 CC protein containing a replacement segment or segments of foreign amino  
 CC acid sequence or sequences grown on a Salmonella, E. coli or  
 CC Enterobacteriaceae host cell, from the host cell and introducing the  
 CC polymer into the animal in conjunction with a carrier or diluent. (1) is  
 CC useful for the expression of recombinant Agfa protein which is useful for  
 CC eliciting an immune response in an animal. In a fimbrial presentation  
 CC system the heterologous antigens are presented in high numbers (up to  
 CC 500,000 copies/cell), the hybrid fimbrial protein possesses both the  
 CC immunogenicity and adhesion properties relevant for an efficient live  
 CC vaccine, the carrier fimbrial subunit proteins are usually strong  
 CC immunogens, which may be important for directing an immune response  
 CC against the inserted epitope, and hybrid fimbriae are easy and  
 CC inexpensive to purify in large amount. The present sequence is given in  
 CC the exemplification of the present invention

XX Sequence 151 AA;

Query Match 73.1%; Score 567; DB 3; Length 151;  
 Best Local Similarity 80.8%; Pred. No. 2.4e-47;  
 Matches 122; Conservative 5; Mismatches 24; Indels 0; Gaps 0;

QY 1 MLLKVAAPAAIVVSGSALAGVVPQGGGNGHNGSGPDSLTSTIYQYGSANAALALQ 60  
 DB 1 MLLKVAAPAAIVVSGSALAGVVPQGGGNGHNGSGPDSLTSTIYQYGSANAALALQ 60

QY 61 SDARKSETTTTQSGYNGADVGADNYDQLVTRVVTHEMAHADQNAKNSDITVQYGG 120

DB 61 SDARKSETTTTQSGYNGADVGADNYDQLVTRVVTHEMAHADQNAKNSDITVQYGG 120

QY 121 NNAALVNQTASDSSVMVRQVFGNNATANQY 151

DB 121 NNAALVNQTASDSSVMVRQVFGNNATANQY 151

RESULT 14

AAB36343  
 ID AAB36343 standard; protein; 151 AA.

XX AAB36343;

XX 26-FEB-2001 (first entry)

XX Escherichia coli CsgA amino acid sequence SEQ ID NO:7.

XX Salmonella; agfa; chromosomal gene replacement; fimbrial; epitope;  
 KW vaccine; immune response; immunogen.

OS Escherichia coli.

XX WO2000060102-A2.

PN 12-OCT-2000.



PF 05-APR-2000; 2000WO-CA000355.  
XX  
PR 05-APR-1999; 99US-0127888P.  
XX  
PA (UYVI-) UNIV VICTORIA.  
XX  
PI White AP, Doran JL, Collison SK, Kay WW;  
XX  
DR WPI; 2000-672631/65.  
DR N-PSDB; AAC64619.  
XX  
XX Recombinant agfa gene having a segment replaced by a foreign DNA sequence  
PT which encodes foreign epitope or antigen, expresses recombinant Agfa  
PT protein useful for eliciting immune response in animal.  
XX  
PS Disclosure; Page 135; 139pp; English.  
XX  
CC The present invention describes a recombinant agfa gene (1) where a  
CC segment of the gene has been replaced by a segment of a foreign DNA  
CC sequence which encodes a foreign epitope or antigen. Also described are:  
CC (1) use of thin aggregative fimbriae (SEF17/TAF) nucleation depended  
CC assembly system of strains of Salmonella, Escherichia coli, and  
CC Enterobacteriaceae for the production of fimbriae comprising recombinant  
CC Agfa, CsgA and Agfa-homologue fimbrial subunits, respectively; (2)  
CC directing recombination of a recombinant gene into the chromosome of the  
CC homologous species; (3) directing recombination of a recombinant gene  
CC back into the chromosome of the homologous species, replacing the native  
CC copy of that gene; and (4) eliciting an immune response in an animal,  
CC comprising separating an amino acid polymer comprising a recombinant Agfa  
CC protein containing a replacement segment or segments of foreign amino  
CC acid sequence or sequences grown on a Salmonella, E. coli or  
CC Enterobacteriaceae host cell, from the host cell and introducing the  
CC polymer into the animal in conjunction with a carrier or diluent. (1) is  
CC useful for the expression of recombinant Agfa protein which is useful for  
CC eliciting an immune response in an animal. In a fimbrial presentation  
CC system the heterologous antigens are presented in high numbers (up to  
CC 500,000 copies/cell), the hybrid fimbrial protein possesses both the  
CC immunogenicity and adhesion properties relevant for an efficient live  
CC vaccine, the carrier fimbrial subunit proteins are usually strong  
CC immunogens, which may be important for directing an immune response  
CC against the inserted epitope, and hybrid fimbriae are easy and  
CC inexpensive to purify in large amount. The present sequence is given in  
CC the exemplification of the present invention  
XX  
SQ Sequence 151 AA;  
Query Match 67.1%; Score 521; DB 3; Length 151;  
Best Local Similarity 69.5%; Pred. No. 7.5e-43;  
Matches 105; Conservative 17; Mismatches 29; Indels 0; Gaps 0;  
QY 1 MKLLKVAAPAAIVVSGSALAGVVPQWGGGNGHGGNSGPDSTLSIYQYGSANAALALQ 60  
DB 1 MKLLKVAIAAIVFSGSALAGVVPQYGGGNGHGGNSGPNSELNIYQYGGNSALALQ 60  
QY 61 SDARKSETTITQSGYNGADYGGADNYDQLVTRVVTHEMAHADOWNAKNSDITVGYGG 120  
DB 61 TDARNSDLTITQGGGNGADYGGQSDSSIDLTORFGNSATLDQWNGKNSMTVKQFGG 120  
QY 121 NNAALVNQTASDSSVMVQVGFNNATANQY 151  
DB 121 GNGAAVDQTASNSVNVTVQVGFNNATAHQY 151  
RESULT 15  
ID ABR82651  
XX ABR82651 standard; protein; 151 AA.  
XX AC ABR82651;  
XX AC  
DT 04-DEC-2003 (first entry)  
XX  
XX E. coli CsgA subunit 15 kDa protein.  
XX

KW Plasma protein; immune response; antibacterial; vaccine; gene therapy.  
XX  
OS Escherichia coli.  
XX  
FN WO2003064446-A2.  
XX  
PD 07-AUG-2003.  
XX  
PF 30-JAN-2003; 2003WO-EP000943.  
XX  
PR 31-JAN-2002; 2002GB-00002275.  
XX  
PA (HANS-) HANSA MEDICAL RES AB.  
XX  
PI Bjoerck L, Olsen A, Wikstroem M, Herwald H;  
XX  
DR WPI; 2003-646136/61.  
DR N-PSDB; ACF36153.  
XX  
XX New isolated peptide capable of binding a mammalian plasma protein,  
PT useful in the manufacture of a medicament for the prevention and/or  
PT treatment of a bacterial infection, such as Escherichia coli, Salmonella  
PT or Shigella infections.  
XX  
PS Disclosure; Page 41-42; 42pp; English.  
XX  
CC The invention relates to an isolated peptide capable of binding a  
CC mammalian plasma protein or of generating an immune response in a mammal  
CC selected from sequences shown in ABR82642, ABR82648-49. The peptide or  
CC antibody is useful for treating a bacterial infection in a human or  
CC animal or in the manufacture of a medicament for the prophylactic  
CC treatment of a bacterial infection, such as Escherichia coli, Salmonella  
CC or Shigella infection. The peptide that is immobilized on a solid support  
CC is also useful as a reagent for determining the ability of a plasma  
CC protein to bind to bacteria. The present sequence represents an E. coli  
CC 15 kDa protein  
XX  
SQ Sequence 151 AA;  
Query Match 66.5%; Score 516; DB 7; Length 151;  
Best Local Similarity 68.9%; Pred. No. 2.3e-42;  
Matches 104; Conservative 17; Mismatches 30; Indels 0; Gaps 0;  
QY 1 MKLLKVAAPAAIVVSGSALAGVVPQWGGGNGHGGNSGPDSTLSIYQYGSANAALALQ 60  
DB 1 MKLLKVEAIAIVFSGSALAGVVPQYGGGNGHGGNSGPNSELNIYQYGGNSALALQ 60  
QY 61 SDARKSETTITQSGYNGADYGGADNYDQLVTRVVTHEMAHADOWNAKNSDITVGYGG 120  
DB 61 TDARNSDLTITQGGGNGADYGGQSDSSIDLTORFGNSATLDQWNGKNSMTVKQFGG 120  
QY 121 NNAALVNQTASDSSVMVQVGFNNATANQY 151  
DB 121 GNGAAVDQTASNSVNVTVQVGFNNATAHQY 151  
Search completed: August 2, 2004, 14:48:27  
Job time : 45.9 secs



GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: August 2, 2004, 14:40:48 ; Search time 12 Seconds  
(without alignments)  
649.627 Million cell updates/sec

Title: US-09-543-407-22  
Perfect score: 776  
Sequence: 1 MKLLKVAFAAIVSGSALA.....DSSVMVROVGFNNATANYQ 151

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 389414 seqs, 51625971 residues

Total number of hits satisfying chosen parameters: 389414

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Issued Patents AA:\*  
1: /cgn2\_6/prodata/2/iaa/5A COMB pep.\*  
2: /cgn2\_6/prodata/2/iaa/5B COMB pep.\*  
3: /cgn2\_6/prodata/2/iaa/6A COMB pep.\*  
4: /cgn2\_6/prodata/2/iaa/6B COMB pep.\*  
5: /cgn2\_6/prodata/2/iaa/6C COMB pep.\*  
6: /cgn2\_6/prodata/2/iaa/backfiles1.pep.\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	677	87.2	151	1	US-08-233-788A-59
2	497	64.0	120	1	US-08-233-788A-57
3	94.5	12.2	738	3	US-08-864-038A-3
4	87	11.2	892	4	US-09-336-447A-5
5	86.5	11.1	238	4	US-09-495-880A-42
6	86.5	11.1	585	4	US-09-620-412C-337
7	86.5	11.1	585	4	US-09-598-419-337
8	86.5	11.1	1752	4	US-09-556-877-180
9	86.5	11.1	1752	4	US-09-620-412C-180
10	86.5	11.1	1752	4	US-09-598-419-180
11	85.5	11.0	589	4	US-09-489-039A-7849
12	82	10.6	975	4	US-09-328-352-4764
13	80	10.3	461	4	US-09-252-991A-24717
14	80	10.3	812	4	US-09-252-991A-30710
15	80	10.3	1034	4	US-09-252-991A-26658
16	79	10.2	943	4	US-09-056-556-204
17	79	10.2	943	4	US-09-072-596-199
18	79	10.2	943	4	US-09-477-135A-131
19	79	10.2	943	4	US-09-072-967-204
20	79	10.2	2315	4	US-05-543-681A-5434
21	78.5	10.1	209	4	US-03-125-619-42
22	78.5	10.1	278	3	US-08-260-883-2
23	78.5	10.1	309	4	US-09-252-991A-22666
24	78.5	10.1	745	4	US-09-336-115C-6
25	78	10.1	392	2	US-08-387-942C-23
26	78	10.1	553	2	US-08-387-942C-2
27	78	10.1	645	4	US-09-919-172-41

28	78	10.1	878	4	US-09-540-236-3401	Sequence 3401, Ap
29	78	10.1	1216	4	US-09-134-000C-5130	Sequence 5130, Ap
30	77.5	10.0	212	4	US-09-125-619-32	Sequence 32, Appl
31	77.5	10.0	437	2	US-08-737-716-2	Sequence 2, Appl
32	77.5	10.0	673	3	US-09-196-387-8	Sequence 8, Appl
33	77.5	10.0	673	4	US-09-841-835-8	Sequence 8, Appl
34	77.5	10.0	702	4	US-09-252-991A-22119	Sequence 22119, A
35	77.5	10.0	941	4	US-09-336-447A-9	Sequence 9, Appl
36	77.5	10.0	949	3	US-09-196-387-10	Sequence 10, Appl
37	77.5	10.0	949	4	US-09-841-835-10	Sequence 10, Appl
38	77.5	10.0	1327	3	US-09-196-387-2	Sequence 2, Appl
39	77.5	10.0	1327	4	US-09-841-835-2	Sequence 8, Appl
40	77.5	10.0	1327	4	US-09-972-115A-8	Sequence 39, Appl
41	77.5	10.0	1690	4	US-09-595-684B-39	Sequence 26, Appl
42	77	9.9	266	4	US-09-435-880A-26	Sequence 32096, A
43	77	9.9	339	4	US-09-252-991A-32096	Sequence 15, Appl
44	77	9.9	873	4	US-09-336-447A-13	Sequence 4, Appl
45	77	9.9	1139	1	US-08-537-210A-4	

ALIGNMENTS

RESULT 1  
US-08-233-788A-59  
; Sequence 59, Application US/08233788A  
; Patent No. 5635617  
; GENERAL INFORMATION:  
; APPLICANT: Doran, James L.  
; APPLICANT: Kay, William W.  
; APPLICANT: Collinson, Karen S.  
; APPLICANT: Clouthier, Sharon C.  
; TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR DETECTION  
; TITLE OF INVENTION: OF SALMONELLA  
; NUMBER OF SEQUENCES: 61  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Seed and Berry  
; STREET: 6300 Columbia Center, 701 Fifth Avenue  
; CITY: Seattle  
; STATE: Washington  
; COUNTRY: U.S.A.  
; ZIP: 98104-7092  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/233,788A  
; FILING DATE: 26-APR-1994  
; CLASSIFICATION: 435  
; ATTORNEY/AGENT INFORMATION:  
; NAME: King, Joshua  
; REGISTRATION NUMBER: 35,570  
; REFERENCE/DOCKET NUMBER: 920043.403C2  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (206) 622-4900  
; TELEFAX: (206) 682-6031  
; TELEX: 3723836 SEEDANBERRY  
; INFORMATION FOR SEQ ID NO: 59:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 151 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
; US-08-233-788A-59

Query Match 87.2%; Score 677; DB 1; Length 151;  
Best Local Similarity 90.1%; Pred. No. 3.3e-59;  
Matches 136; Conservative 2; Mismatches 13; Indels 0; Gaps 0;

1 MKLLKVAFAAIVSGSALAGVVPQWGGNGNGSGPSTLSIYVYGSANALALQ 60

Db 1 MLLKVAFAAIVVSGSALAGVVPQWGGGNGHNGSGSPDSTLSIYQGSANAALALQ 60  
QY 61 SDARKSETTITQSGYNGADVGOGADNYDQLVTRVTHEVAHADQWNAKNSDITVGOYGG 120  
Db 61 SDARKSETTITQSGYNGADVGOGADNSTIELTQNGFRNNATIDQWNAKNSDITVGOYGG 120  
QY 121 NNAALVNTASDSSVMVRQVGFNNATANY 151  
Db 121 NNPALVNTASDSSVMVRQVGFNNATANY 151

RESULT 2  
US-08-233-788A-57  
; Sequence 57, Application US/08233788A  
; Patent No. 5635617  
; GENERAL INFORMATION:  
; APPLICANT: Doran, James L.  
; APPLICANT: Kay, William W.  
; APPLICANT: Collinson, Karen S.  
; APPLICANT: Clouthier, Sharon C.  
; TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR DETECTION  
; TITLE OF INVENTION: OP SALMONELLA  
; NUMBER OF SEQUENCES: 61  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Seed and Berry  
; STREET: 6300 Columbia Center, 701 Fifth Avenue  
; CITY: Seattle  
; STATE: Washington  
; COUNTRY: U.S.A.  
; ZIP: 98104-7092  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patent In Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/233,788A  
; FILING DATE: 26-APR-1994  
; CLASSIFICATION: 435  
; ATTORNEY/AGENT INFORMATION:  
; NAME: King, Joshua  
; REGISTRATION NUMBER: 35,570  
; REFERENCE/DOCKET NUMBER: 920043.403C2  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (206) 622-4900  
; TELEFAX: (206) 682-6031  
; TELEX: 3723836 SEEDANBERRY  
; INFORMATION FOR SEQ ID NO: 57:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 120 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
US-08-233-788A-57

Query Match 64.0%; Score 497; DB 1; Length 120;  
Best Local Similarity 87.5%; Pred. No. 1.2e-41;  
Matches 98; Conservative 2; Mismatches 12; Indels 0; Gaps 0;  
QY 22 VVPQWGGGNGHNGSGSPDSTLSIYQGSANAALALQSDARKSETTITQSGYNGADV 81  
Db 1 VVPQWGGGNGHNGSGSPDSTLSIYQGSANAALALQSDARKSETTITQSGYNGADV 60  
QY 82 GOGADNYDQLVTRVTHEVAHADQWNAKNSDITVGOYGNNAALVNTASDS 133  
Db 61 GOGADNSTIELTQNGFRNNATIDQWNAKNSDITVGOYGNNAALVNTASDS 112

RESULT 3  
US-08-864-038A-3  
; Sequence 3, Application US/08864038A  
; Patent No. 6001592  
; GENERAL INFORMATION:

; APPLICANT: Kunio NAKASHIMA et al.  
; TITLE OF INVENTION: NOVEL POLYPEPTIDE GENE CDNA, VECTOR  
; TITLE OF INVENTION: CONTAINING SAID CDNA, HOST CELLS TRANSFORMED WITH SAID  
; TITLE OF INVENTION: VECTOR, POLYPEPTIDE PRODUCED THEREBY, METHOD OF PRODUCING  
; TITLE OF INVENTION: SAID POLYPEPTIDE, DNA ENCODING SAID POLYPEPTIDE AND ANTIBODY  
; TITLE OF INVENTION: TO SAID POLYPEPTIDE  
; NUMBER OF SEQUENCES: 4  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: 812-5 Hirano  
; STREET: Isehinden  
; CITY: Tsu-city  
; STATE: Mie-prefecture  
; COUNTRY: JAPAN  
; ZIP: 514-01  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Diskette, 3.50 inch, 1.44 MB storage  
; COMPUTER: IBM Compatible  
; OPERATING SYSTEM: Microsoft Windows 95  
; SOFTWARE: Word Perfect 6.1  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/864,038A  
; FILING DATE: May 28, 1997  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: JP 8-184459  
; FILING DATE: 15-July-1996  
; ATTORNEY/AGENT INFORMATION:  
; NAME: C. Bruce Hamburg  
; REGISTRATION NUMBER: 22,389  
; REFERENCE/DOCKET NUMBER: F-5610  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (212)986-2340  
; TELEFAX: (212)953-7733  
; INFORMATION FOR SEQ ID NO: 3:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 738  
; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
; ORIGINAL SOURCE:  
; ORGANISM: Pinctada fucata  
; CELL TYPE: mantle epithelial cell  
; FEATURE:  
; NAME/KEY: peptide  
; LOCATION: from 1 to 738  
; IDENTIFICATION METHOD: E (by experiment)  
US-08-864-038A-3

Query Match 12.2%; Score 94.5; DB 3; Length 738;  
Best Local Similarity 27.6%; Pred. No. 0.48;  
Matches 43; Conservative 14; Mismatches 58; Indels 41; Gaps 6;  
QY 3 LLKVAFAAIVVSGSALAGVVPQWGGGNGHNGSGSPDSTLSIYQGSANAALALQSD 62  
Db 419 LKSSASASASASASAG-----GGGGGNGGNGGGG-----GGAGALA----- 460  
QY 63 ARKSETTITQSGYNGADVGOGADNYDQLVTRVTHEVAHADQWNAKNSDI-TVGOYGN 121  
Db 461 -----AALAAAGAGGGLGGGGGG-----ALAAALAAAGAGGGGFGGLGGLGGGG 507  
QY 122 NAALVNTASDSS-----VWVRQVGFNNATA 148  
Db 508 SAAAAAANAAGGGGRALRRALRRQMRGGGAAA 543

RESULT 4  
US-09-336-447A-5  
; Sequence 5, Application US/09336447A  
; Patent No. 6310190  
; GENERAL INFORMATION:  
; APPLICANT: HANSEN, ERIC J.  
; APPLICANT: ABEI, CHRISTOPH  
; APPLICANT: COPE, LESLIE D.  
; APPLICANT: MACIVER, ISOBEL

APPLICANT: FISKE, MICHAEL J.  
APPLICANT: FREDENBURG, ROSS A.  
TITLE OF INVENTION: USPA1 AND USPA2 ANTIGENS OF MORAXELLA CATARRHALIS  
FILE REFERENCE: AMCY:024  
CURRENT APPLICATION NUMBER: US/09/336,447A  
CURRENT FILING DATE: 1999-06-21  
NUMBER OF SEQ ID NOS: 98  
SOFTWARE: Patentin Ver. 2.1  
SEQ ID NO 5  
LENGTH: 892  
TYPE: PRT  
ORGANISM: Moraxella catarrhalis  
US-09-336-447A-5

Query Match 11.2%; Score 87; DB 4; Length 892;  
Best Local Similarity 26.8%; Pred. No. 3.4;  
Matches 38; Conservative 15; Mismatches 57; Indels 32; Gaps 7;

QY 28 GGGNN-----GGGNSGPDSTLSIYQGSANAALALQSDARKSETTITQSGY--- 75  
DB 89 GGGDYNEAKGNTSTVGGSSNTAKGKSTIGGDTN-----DANGTVSTIGGGYYSRA 141

QY 76 -GNGADVGGADNYDQLVTRVVTHEMAHADQNAKNSDITVQYG---GNNAAIV---N 127  
DB 142 IGDSSSTIGG--YXNQATGKSTVAGGRNNOATGNNSTVAGGSYNQATGNNSTVAGGSHN 199

QY 128 QTASDSSVMVRQVGFNNATAN 149  
DB 200 QATGEGSF--AAGVENKANAN 218

RESULT 5  
US-09-495-880A-42  
Sequence 42, Application US/09495880A  
Patent No. 6667150  
GENERAL INFORMATION:  
APPLICANT: RUDERT, FRITZ  
APPLICANT: GE, LIMING  
APPLICANT: ILAG, VIC  
TITLE OF INVENTION: NOVEL METHOD AND PHAGE FOR THE IDENTIFICATION OF  
TITLE OF INVENTION: NUCLEIC ACID SEQUENCES ENCODING MEMBERS OF A MULTIMERIC  
TITLE OF INVENTION: (POLY)PEPTIDE COMPLEX  
FILE REFERENCE: MORPHO/9  
CURRENT APPLICATION NUMBER: US/09/495,880A  
CURRENT FILING DATE: 2000-02-01  
PRIOR APPLICATION NUMBER: PCT/EP98/04836  
PRIOR FILING DATE: 1998-08-03  
PRIOR APPLICATION NUMBER: EP 97 11 3319.4  
PRIOR FILING DATE: 1997-08-01  
NUMBER OF SEQ ID NOS: 50  
SOFTWARE: Patentin Ver. 2.1  
SEQ ID NO 42  
LENGTH: 238  
TYPE: PRT  
ORGANISM: Artificial Sequence  
FEATURE:  
OTHER INFORMATION: Description of Artificial Sequence: ompA-FLAG-peptide3-  
OTHER INFORMATION: Gene IIs encoded by phage vector fpep3\_LB-Ir3seq (circular)  
US-09-495-880A-42

Query Match 11.1%; Score 86.5; DB 4; Length 238;  
Best Local Similarity 24.3%; Pred. No. 0.69;  
Matches 43; Conservative 24; Mismatches 63; Indels 47; Gaps 8;

QY 5 KVAFAAIVVSGSALAG-----VPPWGGGNNH-NGGNSGPDSTLSI 47  
DB 3 KTAIAVALAGFATVAQADYKVDVCIVYHAHLVAKCGGGGSEFNAGGSGG----- 55

QY 48 YQYSANAALALQSDARKSETTITQSGYNGADYGGADNYDQLVTRVVTHEMAHADQNAKNSDIT 107  
DB 56 ---GSGGSGEGGSGGSGGSEGG--GSGGSGGSGDFYEKANVANKAMTENADE-N 110

QY 108 AKNSDI-----TVQYQG-----GNNAAVLNQT-----SDSSVMVRQVGFNNNA 146

APPLICANT: ALQSDAKGLDVSATDYGAIDGFIGDVSGLANGATGDFAGNSQMAQVGSDGNS 167  
DB 111 ALQSDAKGLDVSATDYGAIDGFIGDVSGLANGATGDFAGNSQMAQVGSDGNS 167

RESULT 6  
US-09-620-412C-337  
Sequence 337, Application US/09620412C  
Patent No. 6448234  
GENERAL INFORMATION:  
APPLICANT: Steven P. Fling  
TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR TREATMENT AND  
TITLE OF INVENTION: DIAGNOSIS OF CHLAMYDIAL INFECTION  
FILE REFERENCE: 210121.469C7  
CURRENT APPLICATION NUMBER: US/09/620,412C  
CURRENT FILING DATE: 2000-07-20  
NUMBER OF SEQ ID NOS: 363  
SOFTWARE: FastSeq for Windows Version 3.0/4.0  
SEQ ID NO 337  
LENGTH: 585  
TYPE: PRT  
ORGANISM: Chlamydia trachomatis  
US-09-620-412C-337

Query Match 11.1%; Score 86.5; DB 4; Length 585;  
Best Local Similarity 31.6%; Pred. No. 2.2;  
Matches 43; Conservative 12; Mismatches 58; Indels 23; Gaps 7;

QY 4 LKVAFAAIVVSGSALAGVPPWGGGNN--HNGGNSGPDSTLSIYQGSANAAL 57  
DB 206 LKAQASAG---NADAWASSPQSGGATTVSDSGSSGSDTSETVPTAKGG----- 257

QY 58 ALQSDARKSETTIT---QSGYNGADYGGADNYDQLVTRVVTHEMAHADQNAKNSDIT 114  
DB 258 GLYTDKNLSITNITGIIELIANNKATDVGGGA-----YVKGTLTCENSHRLOFLKNSSDKQ 312

QY 115 VQG-YGNNAAVLNQT 129  
DB 313 GGGYGEDNITLSNIT 328

RESULT 7  
US-09-598-419-337  
Sequence 337, Application US/09598419  
Patent No. 6565856  
GENERAL INFORMATION:  
APPLICANT: Skeiky, Yasir A.W.  
APPLICANT: Scholler, John  
TITLE OF INVENTION: COMPOUNDS AND METHODS FOR TREATMENT AND  
TITLE OF INVENTION: DIAGNOSIS OF CHLAMYDIAL INFECTION  
FILE REFERENCE: 210121.469C6  
CURRENT APPLICATION NUMBER: US/09/598,419  
CURRENT FILING DATE: 2000-06-20  
NUMBER OF SEQ ID NOS: 357  
SOFTWARE: FastSeq for Windows Version 3.0/4.0  
SEQ ID NO 337  
LENGTH: 585  
TYPE: PRT  
ORGANISM: Chlamydia trachomatis  
US-09-598-419-337

Query Match 11.1%; Score 86.5; DB 4; Length 585;  
Best Local Similarity 31.6%; Pred. No. 2.2;  
Matches 43; Conservative 12; Mismatches 58; Indels 23; Gaps 7;

QY 4 LKVAFAAIVVSGSALAGVPPWGGGNN--HNGGNSGPDSTLSIYQGSANAAL 57  
DB 206 LKAQASAG---NADAWASSPQSGGATTVSDSGSSGSDTSETVPTAKGG----- 257

QY 58 ALQSDARKSETTIT---QSGYNGADYGGADNYDQLVTRVVTHEMAHADQNAKNSDIT 114  
DB 258 GLYTDKNLSITNITGIIELIANNKATDVGGGA-----YVKGTLTCENSHRLOFLKNSSDKQ 312

QY 115 VQG-YGNNAAVLNQT 129

Db 313 GGGIYGEDNITLSNLT 328  
 RESULT 8  
 US-09-556-877-180  
 ; Sequence 180, Application US/09556877  
 ; Patent No. 6432916  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Probst, Peter  
 ; APPLICANT: Bhatia, Ajay  
 ; APPLICANT: Skeiky, Yasir  
 ; APPLICANT: Fling, Steve  
 ; APPLICANT: Maisonneuve, Jeff  
 ; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR TREATMENT AND  
 ; TITLE OF INVENTION: DIAGNOSIS OF CHLAMYDIAL INFECTION  
 ; FILE REFERENCES: 210121.469C5  
 ; CURRENT APPLICATION NUMBER: US/09/556,877  
 ; CURRENT FILING DATE: 2000-04-19  
 ; NUMBER OF SEQ ID NOS: 305  
 ; SOFTWARE: FastSEQ for Windows Version 3.0/4.0  
 ; SEQ ID NO 180  
 ; LENGTH: 1752  
 ; TYPE: PRT  
 ; ORGANISM: Chlamydia  
 US-09-556-877-180

Query Match 11.1%; Score 86.5; DB 4; Length 1752;  
 Best Local Similarity 31.6%; Pred. No. 9;  
 Matches 43; Conservative 12; Mismatches 58; Indels 23; Gaps 7;  
 QY 4 LKVAFAAIVVSGSALAGVVPQWGGGN--HNGGNSGSPDS----TLSTIYQYGSANAAL 57  
 Db 388 LKAQASAG--NADAWASSPQSGGATTVSDSGSSGSDSDTSETVPTAKGG----- 439  
 QY 58 ALQSDARKSETTIT---QSGYGNAGDVGGADNYDQLVTRVTHEMAHADOWNAKNSDIT 114  
 Db 440 GLYTDKNLSITNITGIIETIANNKATDVGGGA-----YVKGTLTLCENSHRLQFLKNSDDQ 494  
 QY 115 VQO-YGGNNAALVNQT 129  
 Db 495 GGGIYGEDNITLSNLT 510

RESULT 9  
 US-09-620-412C-180  
 ; Sequence 180, Application US/09620412C  
 ; Patent No. 6448234  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Steven P. Fling  
 ; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR TREATMENT AND  
 ; TITLE OF INVENTION: DIAGNOSIS OF CHLAMYDIAL INFECTION  
 ; FILE REFERENCES: 210121.469C7  
 ; CURRENT APPLICATION NUMBER: US/09/620,412C  
 ; CURRENT FILING DATE: 2000-07-20  
 ; NUMBER OF SEQ ID NOS: 363  
 ; SOFTWARE: FastSEQ for Windows Version 3.0/4.0  
 ; SEQ ID NO 180  
 ; LENGTH: 1752  
 ; TYPE: PRT  
 ; ORGANISM: Chlamydia  
 US-09-620-412C-180

Query Match 11.1%; Score 86.5; DB 4; Length 1752;  
 Best Local Similarity 31.6%; Pred. No. 9;  
 Matches 43; Conservative 12; Mismatches 58; Indels 23; Gaps 7;  
 QY 4 LKVAFAAIVVSGSALAGVVPQWGGGN--HNGGNSGSPDS----TLSTIYQYGSANAAL 57  
 Db 388 LKAQASAG--NADAWASSPQSGGATTVSDSGSSGSDSDTSETVPTAKGG----- 439  
 QY 58 ALQSDARKSETTIT---QSGYGNAGDVGGADNYDQLVTRVTHEMAHADOWNAKNSDIT 114

Db 440 GLYTDKNLSITNITGIIETIANNKATDVGGGA-----YVKGTLTLCENSHRLQFLKNSDDQ 494  
 QY 115 VQO-YGGNNAALVNQT 129  
 Db 495 GGGIYGEDNITLSNLT 510  
 RESULT 10  
 US-09-598-419-180  
 ; Sequence 180, Application US/09598419  
 ; Patent No. 6565856  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Skeiky, Yasir A.W.  
 ; APPLICANT: Scholler, John  
 ; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR TREATMENT AND  
 ; TITLE OF INVENTION: DIAGNOSIS OF CHLAMYDIAL INFECTION  
 ; FILE REFERENCES: 210121.469C6  
 ; CURRENT APPLICATION NUMBER: US/09/598,419  
 ; CURRENT FILING DATE: 2000-06-20  
 ; NUMBER OF SEQ ID NOS: 357  
 ; SOFTWARE: FastSEQ for Windows Version 3.0/4.0  
 ; SEQ ID NO 180  
 ; LENGTH: 1752  
 ; TYPE: PRT  
 ; ORGANISM: Chlamydia  
 US-09-598-419-180

Query Match 11.1%; Score 86.5; DB 4; Length 1752;  
 Best Local Similarity 31.6%; Pred. No. 9;  
 Matches 43; Conservative 12; Mismatches 58; Indels 23; Gaps 7;  
 QY 4 LKVAFAAIVVSGSALAGVVPQWGGGN--HNGGNSGSPDS----TLSTIYQYGSANAAL 57  
 Db 388 LKAQASAG--NADAWASSPQSGGATTVSDSGSSGSDSDTSETVPTAKGG----- 439  
 QY 58 ALQSDARKSETTIT---QSGYGNAGDVGGADNYDQLVTRVTHEMAHADOWNAKNSDIT 114  
 Db 440 GLYTDKNLSITNITGIIETIANNKATDVGGGA-----YVKGTLTLCENSHRLQFLKNSDDQ 494  
 QY 115 VQO-YGGNNAALVNQT 129  
 Db 495 GGGIYGEDNITLSNLT 510

RESULT 11  
 US-09-489-039A-7849  
 ; Sequence 7849, Application US/09489039A  
 ; Patent No. 6610836  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Gary Breton et. al  
 ; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO KLEBSIELLA  
 ; TITLE OF INVENTION: PNEUMONIAE FOR DIAGNOSTICS AND THERAPEUTICS  
 ; FILE REFERENCES: 2709.2004001  
 ; CURRENT APPLICATION NUMBER: US/09/489,039A  
 ; CURRENT FILING DATE: 2000-01-27  
 ; PRIOR APPLICATION NUMBER: US 60/117,747  
 ; PRIOR FILING DATE: 1999-01-29  
 ; NUMBER OF SEQ ID NOS: 14342  
 ; SEQ ID NO 7849  
 ; LENGTH: 589  
 ; TYPE: PRT  
 ; ORGANISM: Klebsiella pneumoniae  
 US-09-489-039A-7849

Query Match 11.0%; Score 85.5; DB 4; Length 589;  
 Best Local Similarity 23.2%; Pred. No. 2.8;  
 Matches 32; Conservative 17; Mismatches 68; Indels 21; Gaps 3;  
 QY 29 GGNHNGGNSGSPDSTLSIYQYGS-----ANAAALQSDARKSETTITQSGYGNAGD 80  
 Db 259 GGLDRNGANANGQDTDFGIYAPDTLTLTERIEINGLRLDNYHTKYDSATACGSGRGAI 318  
 QY 81 VQOAGADNYDQLVTRVTHEMAHADOWNA-----KNSDITVQYGGNNAALVNQTASD 133

Db 319 ACPPGOSTSPVTTVDIAKSGNLVWAKALYRLTQGNVYV-----NYAISQPPGGS 372  
QY 134 SVMVQVGFQGNATANQY 151  
Db 373 SFALAASGSGSANRTDF 390

## RESULT 12

US-09-328-352-4764  
; Sequence 4764, Application US/09328352  
; Patent No. 6562958  
; GENERAL INFORMATION:  
; APPLICANT: Gary L. Breton et al.  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO ACINETOBACTER  
; TITLE OF INVENTION: BAUMANNII FOR DIAGNOSTICS AND THERAPEUTICS  
; FILE REFERENCE: GTC99-03PA  
; CURRENT APPLICATION NUMBER: US/09/328,352  
; CURRENT FILING DATE: 1999-06-04  
; NUMBER OF SEQ ID NOS: 8252  
; SEQ ID NO 4764  
; LENGTH: 975  
; TYPE: PRT  
; ORGANISM: Acinetobacter baumannii  
US-09-328-352-4764

Query Match 10.6%; Score 82; DB 4; Length 975;  
Best Local Similarity 24.8%; Pred. No. 12; Mismatches 47; Indels 58; Gaps 7;  
Matches 38; Conservative 15;

QY 15 SGLAGVVPQWGGGNGHNG-GNSGPDSTLSIYQY-----SANA 55  
Db 300 AGNGIA-----SGNGHNYGIGNGGDDVDITAPITGVNLISGNSFTLIGNSSSSSVNT 353  
QY 56 ALALQSDARKSETI-----TQSGYG-----NGADVGGADNYDQLVTRVVTHEM 100  
Db 354 APTTNTVNDNTIDNGSGGTGSGSGSGGGLNGAASNGEHNIG-----402  
QY 101 AHADQWAKNSDIT-----VGQYGGNNAALVNQTSDS 133  
Db 403 --TGNGGDDVDITSPITGIFNFGNSFSLIGNSSSSS 438

## RESULT 13

US-09-252-991A-24717  
; Sequence 24717, Application US/09252991A  
; Patent No. 6551795  
; GENERAL INFORMATION:  
; APPLICANT: Marc J. Rubenfield et al.  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS  
; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS  
; FILE REFERENCE: 107196.136  
; CURRENT APPLICATION NUMBER: US/09/252,991A  
; CURRENT FILING DATE: 1999-02-18  
; PRIOR APPLICATION NUMBER: US 60/074,788  
; PRIOR FILING DATE: 1998-02-18  
; PRIOR FILING DATE: 1998-02-18  
; PRIOR FILING DATE: 1998-07-27  
; NUMBER OF SEQ ID NOS: 33142  
; SEQ ID NO 24717  
; LENGTH: 461  
; TYPE: PRT  
; ORGANISM: Pseudomonas aeruginosa  
US-09-252-991A-24717

Query Match 10.3%; Score 80; DB 4; Length 461;  
Best Local Similarity 25.6%; Pred. No. 7;  
Matches 43; Conservative 20; Mismatches 67; Indels 38; Gaps 8;

QY 6 VAAPFAIIVSG-SALAGVVPQWGGGNGHNGGKSSGPDSTLSIYQYGSANA---ALALQS 61  
Db 300 VAALPEVARSGPAPSGTAPAGGAA---GKSPAGLRLARRSSFPSSSATPPATTFY 356

QY 62 DARKSETTITQSG-YNGA-----DVGGADNYDQLVT-----RVVTHEMAHADOWN- 107  
Db 357 AARAPMAITEPGRSGTGADPRRTAGTGEQGLVVKDFAHPAYRLVTOEIEGDSWNG 416  
QY 108 -----AKNSDITVQYGGNNAALVNQTSDSVVMVQVGFQGNNA 146  
Db 417 DVFLRIDMAAELOGDFMGDIADGNOAVVASQVISET-----VGNHA 456

## RESULT 14

US-09-252-991A-30710  
; Sequence 30710, Application US/09252991A  
; Patent No. 6551795  
; GENERAL INFORMATION:  
; APPLICANT: Marc J. Rubenfield et al.  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS  
; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS  
; FILE REFERENCE: 107196.136  
; CURRENT APPLICATION NUMBER: US/09/252,991A  
; CURRENT FILING DATE: 1999-02-18  
; PRIOR APPLICATION NUMBER: US 60/074,788  
; PRIOR FILING DATE: 1998-02-18  
; PRIOR APPLICATION NUMBER: US 60/094,190  
; PRIOR FILING DATE: 1998-07-27  
; NUMBER OF SEQ ID NOS: 33142  
; SEQ ID NO 30710  
; LENGTH: 812  
; TYPE: PRT  
; ORGANISM: Pseudomonas aeruginosa  
US-09-252-991A-30710

Query Match 10.3%; Score 80; DB 4; Length 812;  
Best Local Similarity 26.7%; Pred. No. 15;  
Matches 31; Conservative 10; Mismatches 39; Indels 36; Gaps 5;

QY 51 GSANAALALQSDARKSE-----TTTQSGYGGNGADVGGADNYDQLVTRVVTHEMAHAD 104  
Db 550 GLLNIGVTRDSDRRYSERVWRSRSTPSQGLGNLGYGGGASRYQQ-----AD 597  
QY 105 -QWNAKNSDITVQYVG--GN-----NAALVNQTSDSVVMVQVGF 142  
Db 598 LTRWQNVQLQGLYGETGNYTRWADLSGSLVWMDNAVFNASNRINDAFVLVSTKGY 653

## RESULT 15

US-09-252-991A-26658  
; Sequence 26658, Application US/09252991A  
; Patent No. 6551795  
; GENERAL INFORMATION:  
; APPLICANT: Marc J. Rubenfield et al.  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS  
; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS  
; FILE REFERENCE: 107196.136  
; CURRENT APPLICATION NUMBER: US/09/252,991A  
; CURRENT FILING DATE: 1999-02-18  
; PRIOR APPLICATION NUMBER: US 60/074,788  
; PRIOR FILING DATE: 1998-02-18  
; PRIOR APPLICATION NUMBER: US 60/094,190  
; PRIOR FILING DATE: 1998-07-27  
; NUMBER OF SEQ ID NOS: 33142  
; SEQ ID NO 26658  
; LENGTH: 1034  
; TYPE: PRT  
; ORGANISM: Pseudomonas aeruginosa  
US-09-252-991A-26658

Query Match 10.3%; Score 80; DB 4; Length 1034;  
Best Local Similarity 26.7%; Pred. No. 20;  
Matches 47; Conservative 19; Mismatches 62; Indels 48; Gaps 10;

QY 14 VSGSALAGVVPQWGGGNGHNG-----GNSGPDSTLSIYQYGS-ANAALALQSDAR-- 64  
Db 748 VSDSASGGV-BAGGAGNTGGLVGLSSGGEIFRSQASGSVYKGGGLATGLIGKAENGGM 806

```
QY 65 -----KSETTITQSGYGNQADVGQADNYDQV-----TRV----- 95
Db 807 LGNLKASGSVTDQG---GADLGLVGNNSQSAIETABATGKVSQSGNSRVGGLIGHNLGG 863
QY 96 -VTHEMAHADQWNAKNSDITVQYGGNNAALVNOTASD--SSVMVRQVG--FGNNA 146
Db 864 SVAHAI SRGDSVGGFNS-LVGGLVGHGGLVNVVDASGRVSAASASVGGLVGSNA 918
```

Search completed: August 2, 2004, 14:58:35  
Job time : 13 secs



GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: August 2, 2004, 14:54:48 ; Search time 36.8 Seconds  
(without alignments)  
1287.123 Million cell updates/sec

Title: US-09-543-407-22  
Perfect score: 776  
Sequence: 1 MKLLKVAFAAIVWSSGALA.....DSSVMVROVGFNNATANQY 151

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 1291235 seqs, 313682936 residues

Total number of hits satisfying chosen parameters: 1291235

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000  
Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

- Database : Published Applications AA.\*
- 1: /cgn2\_6/ptodata/2/pubpaa/US07\_PUBCOMB.pep.\*
  - 2: /cgn2\_6/ptodata/2/pubpaa/PCT\_NEW\_PUB.pep.\*
  - 3: /cgn2\_6/ptodata/2/pubpaa/US06\_NEW\_PUB.pep.\*
  - 4: /cgn2\_6/ptodata/2/pubpaa/US06\_PUBCOMB.pep.\*
  - 5: /cgn2\_6/ptodata/2/pubpaa/US07\_NEW\_PUB.pep.\*
  - 6: /cgn2\_6/ptodata/2/pubpaa/PCTUS\_PUBCOMB.pep.\*
  - 7: /cgn2\_6/ptodata/2/pubpaa/US08\_NEW\_PUB.pep.\*
  - 8: /cgn2\_6/ptodata/2/pubpaa/US08\_PUBCOMB.pep.\*
  - 9: /cgn2\_6/ptodata/2/pubpaa/US09\_PUBCOMB.pep.\*
  - 10: /cgn2\_6/ptodata/2/pubpaa/US09\_PUBCOMB.pep.\*
  - 11: /cgn2\_6/ptodata/2/pubpaa/US09\_NEW\_PUB.pep.\*
  - 12: /cgn2\_6/ptodata/2/pubpaa/US10\_PUBCOMB.pep.\*
  - 13: /cgn2\_6/ptodata/2/pubpaa/US10\_PUBCOMB.pep.\*
  - 14: /cgn2\_6/ptodata/2/pubpaa/US10\_PUBCOMB.pep.\*
  - 15: /cgn2\_6/ptodata/2/pubpaa/US10\_PUBCOMB.pep.\*
  - 16: /cgn2\_6/ptodata/2/pubpaa/US10\_NEW\_PUB.pep.\*
  - 17: /cgn2\_6/ptodata/2/pubpaa/US60\_NEW\_PUB.pep.\*
  - 18: /cgn2\_6/ptodata/2/pubpaa/US60\_PUBCOMB.pep.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Query Match	Length	ID	Description
1	518	66.8	151 12	US-09-741-873B-4
2	518	66.8	151 12	US-09-741-873B-4
3	440	56.7	131 12	US-09-741-873B-2
4	440	56.7	131 12	US-09-741-873B-2
5	95	12.2	445 15	US-10-369-493-20638
6	95	12.2	597 9	US-09-793-306-146
7	94	12.1	271 16	US-10-437-963-147343
8	93	12.0	369 12	US-10-425-114-56041
9	93	12.0	486 12	US-10-424-599-275468
10	93	12.0	507 12	US-10-425-114-57763
11	92.5	11.9	251 12	US-09-880-748-1130
12	92.5	11.9	251 12	US-10-293-418-1130
13	91	11.7	1751 9	US-09-841-132-445
14	91	11.7	1751 9	US-09-841-132-594
15	91	11.7	1751 16	US-10-467-534-45

16	89.5	11.5	251	10	US-09-880-748-1122	Sequence 1122, Ap
17	89.5	11.5	251	12	US-10-293-418-1122	Sequence 1122, Ap
18	89	11.5	193	16	US-10-437-963-148500	Sequence 148500,
19	88.5	11.4	145	16	US-10-437-963-147748	Sequence 147748,
20	88.5	11.4	197	12	US-10-425-114-67750	Sequence 67750, A
21	88	11.3	503	16	US-10-437-963-158876	Sequence 158876,
22	87.5	11.3	154	16	US-10-437-963-162284	Sequence 162284,
23	87.5	11.3	204	12	US-10-424-599-203972	Sequence 203972,
24	87.5	11.3	486	15	US-10-369-493-20619	Sequence 20619, A
25	87.5	11.3	689	16	US-10-437-963-188971	Sequence 188971,
26	87	11.2	892	10	US-09-952-267-5	Sequence 5, Appli
27	86.5	11.1	191	16	US-10-437-963-105413	Sequence 105413,
28	86.5	11.1	238	12	US-10-634-862-42	Sequence 42, Appl
29	86.5	11.1	585	9	US-09-841-132-337	Sequence 337, App
30	86.5	11.1	1752	9	US-09-841-132-180	Sequence 180, App
31	86	11.1	448	16	US-10-437-963-140685	Sequence 140685,
32	86	11.1	735	12	US-10-425-114-63965	Sequence 63965, A
33	86	11.1	6310	12	US-10-282-122A-67793	Sequence 67793, A
34	85.5	11.0	270	16	US-10-437-963-122263	Sequence 122263,
35	85.5	11.0	580	12	US-10-647-057-4	Sequence 4, Appli
36	85	11.0	354	10	US-09-820-843A-21	Sequence 21, Appl
37	85	11.0	400	16	US-10-437-963-186417	Sequence 186417,
38	84.5	10.9	189	16	US-10-437-963-170736	Sequence 170736,
39	84.5	10.9	1276	16	US-10-437-963-168952	Sequence 168952,
40	84	10.8	242	12	US-10-425-114-61520	Sequence 61520, A
41	84	10.8	253	10	US-09-880-748-2098	Sequence 2098, Ap
42	84	10.8	253	12	US-10-293-418-2098	Sequence 2098, Ap
43	84	10.8	255	10	US-09-880-748-1153	Sequence 1153, Ap
44	84	10.8	255	12	US-10-293-418-1153	Sequence 1153, Ap
45	84	10.8	1448	16	US-10-408-765A-998	Sequence 998, App

ALIGNMENTS

RESULT 1  
US-09-741-873B-4  
; Sequence 4, Application US/09741873B  
; Publication No. US20020081722A1  
; GENERAL INFORMATION:  
; APPLICANT: Normark, Staffan  
; APPLICANT: Olsen, Arne  
; TITLE OF INVENTION: Fibronectin Binding Protein As Well As Its Preparation  
; FILE REFERENCE: 012889-084  
; CURRENT APPLICATION NUMBER: US/09/741.873B  
; CURRENT FILING DATE: 2003-04-04  
; PRIOR APPLICATION NUMBER: SE 8801723-1  
; PRIOR FILING DATE: 1998-05-06  
; PRIOR APPLICATION NUMBER: US 08/978,878  
; PRIOR FILING DATE: 1997-11-26  
; PRIOR APPLICATION NUMBER: US 07/347,189  
; PRIOR FILING DATE: 1989-05-04  
; PRIOR APPLICATION NUMBER: US 07/789,437  
; PRIOR FILING DATE: 1991-11-06  
; PRIOR APPLICATION NUMBER: US 07/970,846  
; PRIOR FILING DATE: 1992-11-03  
; PRIOR APPLICATION NUMBER: US 08/187,865  
; PRIOR FILING DATE: 1994-01-28  
; PRIOR APPLICATION NUMBER: US 08/318,519  
; NUMBER OF SEQ ID NOS: 10  
; SOFTWARE: PatentIn version 3.0  
; SEQ ID NO 4  
; LENGTH: 151  
; TYPE: PRT  
; ORGANISM: Escherichia coli  
US-09-741-873B-4

Query Match 66.8%; Score 518; DB 12; Length 151;  
Best Local Similarity 68.9%; Pred. No. 1.2e-44;  
Matches 104; Conservative 18; Mismatches 29; Indels 0; Gaps 0;  
QY 1 MKLLKVAFAAIVWSSGALAGVVPQWGGGNGHNGSSGPDSTLSIYQYGSANAALALQ 60

```

Db 1 MKLLKVAALAAIYVSGSAGVVPQYGGGNGHGGGNSGPNSELNIYQYGGNSALALQ 60
    :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 61 SDARKSETTITQSGYNGADVGAGADNYDLVTRVVTHEMAHADOWNAKNSDITVQYGG 120
    :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 61 TDARNSDLTITQGGNGADVGAGADNYDLVTRVVTHEMAHADOWNAKNSDITVQYGG 120
    :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 121 NNAALVNOTASDSSVMVRQYGFNNATANQY 151
    :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 121 NGGAADVDTASNSVNVTVQYGFNNATANQY 151
    :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:

```

```

RESULT 2
US-09-741-873B-4
; Sequence 4, Application US/09741873B
; Publication No. US2004009695A9
; GENERAL INFORMATION:
; APPLICANT: Normark, Staffan
; APPLICANT: Olsen, Arne
; TITLE OF INVENTION: Fibronectin Binding Protein As Well As Its Preparation
; FILE REFERENCE: 012889-084
; CURRENT APPLICATION NUMBER: US/09/741,873B
; CURRENT FILING DATE: 2003-04-04
; PRIOR APPLICATION NUMBER: SE 8801723-1
; PRIOR FILING DATE: 1998-05-06
; PRIOR APPLICATION NUMBER: US 08/978,878
; PRIOR FILING DATE: 1997-11-26
; PRIOR APPLICATION NUMBER: US 07/347,189
; PRIOR FILING DATE: 1999-05-04
; PRIOR APPLICATION NUMBER: US 07/789,437
; PRIOR FILING DATE: 1991-11-06
; PRIOR APPLICATION NUMBER: US 07/970,846
; PRIOR FILING DATE: 1992-11-03
; PRIOR APPLICATION NUMBER: US 08/187,865
; PRIOR FILING DATE: 1994-01-28
; PRIOR APPLICATION NUMBER: US 08/318,519
; PRIOR FILING DATE: 1994-10-05
; NUMBER OF SEQ ID NOS: 10
; SOFTWARE: Patentin version 3.0
; SEQ ID NO 4
; LENGTH: 151
; TYPE: PRT
; ORGANISM: Escherichia coli
US-09-741-873B-4

```

```

Query Match 66.8%; Score 518; DB 12; Length 151;
Best Local Similarity 68.9%; Pred. No. 1.2e-44;
Matches 104; Conservative 18; Mismatches 29; Indels 0; Gaps 0;

QY 1 MKLLKVAALAAIYVSGSAGVVPQYGGGNGHGGGNSGPNSELNIYQYGGNSALALQ 60
    :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 1 MKLLKVAALAAIYVSGSAGVVPQYGGGNGHGGGNSGPNSELNIYQYGGNSALALQ 60
    :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 61 SDARKSETTITQSGYNGADVGAGADNYDLVTRVVTHEMAHADOWNAKNSDITVQYGG 120
    :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 61 TDARNSDLTITQGGNGADVGAGADNYDLVTRVVTHEMAHADOWNAKNSDITVQYGG 120
    :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 121 NNAALVNOTASDSSVMVRQYGFNNATANQY 151
    :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 121 NGGAADVDTASNSVNVTVQYGFNNATANQY 151
    :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:

```

```

RESULT 3
US-09-741-873B-2
; Sequence 2, Application US/09741873B
; Publication No. US20020081722A1
; GENERAL INFORMATION:
; APPLICANT: Normark, Staffan
; APPLICANT: Olsen, Arne
; TITLE OF INVENTION: Fibronectin Binding Protein As Well As Its Preparation
; FILE REFERENCE: 012889-084
; CURRENT APPLICATION NUMBER: US/09/741,873B
; CURRENT FILING DATE: 2003-04-04

```

```

; PRIOR APPLICATION NUMBER: SE 8801723-1
; PRIOR FILING DATE: 1998-05-06
; PRIOR APPLICATION NUMBER: US 08/978,878
; PRIOR FILING DATE: 1997-11-26
; PRIOR APPLICATION NUMBER: US 07/347,189
; PRIOR FILING DATE: 1989-05-04
; PRIOR APPLICATION NUMBER: US 07/789,437
; PRIOR FILING DATE: 1991-11-06
; PRIOR APPLICATION NUMBER: US 07/970,846
; PRIOR FILING DATE: 1992-11-03
; PRIOR APPLICATION NUMBER: US 08/187,865
; PRIOR FILING DATE: 1994-01-28
; PRIOR APPLICATION NUMBER: US 08/318,519
; PRIOR FILING DATE: 1994-10-05
; NUMBER OF SEQ ID NOS: 10
; SOFTWARE: Patentin version 3.0
; SEQ ID NO 2
; LENGTH: 131
; TYPE: PRT
; ORGANISM: Escherichia coli
US-09-741-873B-2

Query Match 56.7%; Score 440; DB 12; Length 131;
Best Local Similarity 65.6%; Pred. No. 8.3e-37;
Matches 86; Conservative 17; Mismatches 28; Indels 0; Gaps 0;

QY 21 GWVPQWGGGNGHGGGNSGPNSELNIYQYGGNSALALQSDARKSETTITQSGYNGAD 80
    :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 1 GWVPQYGGGNGHGGGNSGPNSELNIYQYGGNSALALQTDARNSDLTITQGGNGAD 60
    :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 81 VGGADNYDLVTRVVTHEMAHADOWNAKNSDITVQYGGNSALALVNOTASDSSVMVRQV 140
    :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 61 VGGQSDSSDLTQYGFNGSATLDQWNGKNSMTVQYGGNGAAVDQYASNSVNVTVQV 120
    :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 141 GFGNNATANQY 151
    :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 121 GFGNNATANQY 131
    :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:

```

```

RESULT 4
US-09-741-873B-2
; Sequence 2, Application US/09741873B
; Publication No. US2004009695A9
; GENERAL INFORMATION:
; APPLICANT: Normark, Staffan
; APPLICANT: Olsen, Arne
; TITLE OF INVENTION: Fibronectin Binding Protein As Well As Its Preparation
; FILE REFERENCE: 012889-084
; CURRENT APPLICATION NUMBER: US/09/741,873B
; CURRENT FILING DATE: 2003-04-04
; PRIOR APPLICATION NUMBER: SE 8801723-1
; PRIOR FILING DATE: 1998-05-06
; PRIOR APPLICATION NUMBER: US 08/978,878
; PRIOR FILING DATE: 1997-11-26
; PRIOR APPLICATION NUMBER: US 07/347,189
; PRIOR FILING DATE: 1989-05-04
; PRIOR APPLICATION NUMBER: US 07/789,437
; PRIOR FILING DATE: 1991-11-06
; PRIOR APPLICATION NUMBER: US 07/970,846
; PRIOR FILING DATE: 1992-11-03
; PRIOR APPLICATION NUMBER: US 08/187,865
; PRIOR FILING DATE: 1994-01-28
; PRIOR APPLICATION NUMBER: US 08/318,519
; PRIOR FILING DATE: 1994-10-05
; NUMBER OF SEQ ID NOS: 10
; SOFTWARE: Patentin version 3.0
; SEQ ID NO 2
; LENGTH: 131
; TYPE: PRT
; ORGANISM: Escherichia coli
US-09-741-873B-2

```

```

Query Match 56.7%; Score 440; DB 12; Length 131;

```

[illegible]

RESULT 5  
 US-10-369-493-20638  
 ; Sequence 20638, Application US/10369493  
 ; Publication No. US20030233675A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Cao, Yongwei  
 ; APPLICANT: Hinkle, Gregory J.  
 ; APPLICANT: Slater, Steven C.  
 ; APPLICANT: Goldman, Barry S.  
 ; APPLICANT: Chen, Xianfeng  
 ; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF  
 ; TITLE OF INVENTION: PLANTS WITH IMPROVED PROPERTIES  
 ; FILE REFERENCE: 38-10(52052)B  
 ; CURRENT APPLICATION NUMBER: US/10/369,493  
 ; CURRENT FILING DATE: 2003-02-28  
 ; PRIOR APPLICATION NUMBER: US 60/360,039  
 ; PRIOR FILING DATE: 2002-02-21  
 ; NUMBER OF SEQ ID NOS: 47374  
 ; SEQ ID NO 20638  
 ; LENGTH: 445  
 ; TYPE: PRT  
 ; ORGANISM: Rhodopseudomonas palustris  
 ; FEATURE:  
 ; NAME/KEY: unsure  
 ; LOCATION: (1)..(445)  
 ; OTHER INFORMATION: unsure at all Xaa locations  
 US-10-369-493-20638

Query Match	12.2%;	Score 95;	DB 15;	Length 445;
Best Local Similarity	24.6%;	Pred. N.O. 41;		
Matches 35;	Conservative 22;	Mismatches 59;	Indels 25;	Gaps 5
Qy	30	GNHNGGGNSS-----GPDSTLSIYQVGSANAA-----LALQSDARKSET	68	
Db	89	GKQSGAGNSAAIFQSGTSDVELQQTGTSGAVPSGKWNNDPGVFNKITQDSSNGSKV	148	
Qy	69	TITQSGYGNADVGGCQADNYDQLVTRVVTVMHAHADQW-NAKNSDITTVQQ-----YGGN-NA	123	
Db	149	SVIQDGKKNVPSIKQGTNGTNSVNOQIGEWGNAYVYRGIGAEETDASTGNALPTGGINV	208	
Qy	124	ALVNQATSDSSVMVTRQVFGNN	145	
Db	209	ASITQNSAGLNAYAVAVOGGNS	230	

RESULT 6  
US-09-793-306-146  
; Sequence 146, Application US/09793306  
; Patent No. US20020098200A1  
; GENERAL INFORMATION:  
; APPLICANT: Campos-Neto, Antonio  
; APPLICANT: Skeiky, Yasir  
; APPLICANT: Ovendale, Pamela  
; APPLICANT: Jen, Shyian  
; APPLICANT: Lodes, Michael  
; APPLICANT: Corixa Corporation  
; TITLE OF INVENTION: Compounds and Methods for Diagnosis and Immunotherapy

```

; TITLE OF INVENTION: of Tuberculosis
; FILE REFERENCE: 014058-008740US
; CURRENT APPLICATION NUMBER: US/09/793,306
; CURRENT FILING DATE: 2001-02-26
; PRIOR APPLICATION NUMBER: US 60/185,037
; PRIOR FILING DATE: 2000-02-25
; PRIOR APPLICATION NUMBER: US 60/223,828
; PRIOR FILING DATE: 2000-08-08
; NUMBER OF SEQ ID NOS: 164
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 146
; LENGTH: 597
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence:mTTC#3-His
US-09-793-306-146

```

Query Match	12.2%	Score 95;	DB 9;	Length 597;	
Best Local Similarity	27.4%;	Pred. No. 0.59;			
Matches	34;	Conservative 14;	Mismatches 50;	Indels 26;	Gaps 4
Qy	26	WCGGNGNHGGNSGPDSTLSIIYQGSANAALALQSDARKSETITITQSGVGNAGDVGQCA	85		
Db	358	FGNSGNNITGFNSG-NNNVGFFNSGNN-----FQFGNAGDINTGTF	398		
Qy	86	DNYDLQVTRVVTHEMAHADOWNAKNSDITVQYGGNNAALVNQTSADSSVMYRVQVGFGNN	145		
Db	399	GNAGDTNTGFGNAGFGNMGIGNAGNEDMGVNGGSFNVGVGN--AGNQS-----VGFGNA	451		
Qy	146	ATAN	149		
Db	452	GTLN	455		

```

RESULT 7
US-10-437-963-147343
; Sequence 147343, Application US/10437963
; Publication No. US20040123343A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J.
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; APPLICANT: Wu, Wei
; APPLICANT: Boukharov, Andrey A.
; APPLICANT: Barbazuk, Brad
; APPLICANT: Li, Ping
; TITLE OF INVENTION: Rice Nucleic Acid Molecules and Other Molecules Associated With
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(53221)B
; CURRENT APPLICATION NUMBER: US/10/437,963
; CURRENT FILING DATE: 2003-05-14
; NUMBER OF SEQ ID NOS: 204966
; SEQ ID NO 147343
; LENGTH: 271
; TYPE: PRT
; ORGANISM: Oryza sativa
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MFT4530_47881C.1.psp
US-10-437-963-147343

```

Query Match	12.1%	Score 94;	DB 16;	Length 271;
Best Local Similarity	26.2%	Pred. No. 0.28;		
Matches	37;	Conservative	19;	Mismatches 67;
			Indels	18;
			Gaps	6

  

QY	14	VSGSALAGVFPQWGGG	-GNHNGGNSGPDFSTL	SIYQGSANAALALOSDARKSETTTT	QY	72
		:	:	:		:
		:	:	:		:
DB	7	VVGRAAAAAYPEVSGDGGSGGGEGGE	-----GGDGSVAANVPEAGSGDGRSS			58
		:	:	:		:
		:	:	:		:
QY	73	SGYNGNGADVGGCAGNYDQLVTRVVVTH	MAHADQWNAKNSDITV	GOYCGNNAALVQNTASD		132
		:	:	:		:
		:	:	:		:
DB	59	CGEGGGSGGGLGRQRYWNSRLSTER	QLVDH-VFKNSDVVCDVFSG	-VGFIAIS		112
		:	:	:		:
		:	:	:		:

QY 133 SSVMVQVGFQGN--NATANQY 151  
 Db 113 AARKVIV-YANDLNPATVEY 132

## RESULT 8

US-10-425-114-56041  
 ; Sequence 56041, Application US/10425114  
 ; Publication No. US20040034888A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Liu, Jingdong  
 ; APPLICANT: Zhou, Yihua  
 ; APPLICANT: Kovalic, David K.  
 ; APPLICANT: Screen, Steven E  
 ; APPLICANT: Tabaska, Jack E  
 ; APPLICANT: Cao, Yongwei  
 ; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With  
 ; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement  
 ; FILE REFERENCE: 38-21(53313)B  
 ; CURRENT APPLICATION NUMBER: US/10/425,114  
 ; CURRENT FILING DATE: 2003-04-28  
 ; NUMBER OF SEQ ID NOS: 73128  
 ; SEQ ID NO 56041  
 ; LENGTH: 369  
 ; TYPE: PRT  
 ; ORGANISM: Glycine max  
 ; FEATURE:  
 ; OTHER INFORMATION: Clone ID: 701205720\_FLI.pap  
 US-10-425-114-56041

Query Match 12.0%; Score 93; DB 12; Length 369;  
 Best Local Similarity 26.3%; Pred. No. 0.51;  
 Matches 36; Conservative 18; Mismatches 63; Indels 20; Gaps 6;

QY 14 VSGSALAGVVPQGGGNGHNGGNSGPDSTLSIYQVGSANAALQSDARKSETTITQS 73  
 Db 13 VRSSASKG-KPPLSAGNNNGGNDWDPD-----GFGSARGG-----ADLRNQSTGDYR 62  
 QY 74 GYNGADVGGADNYDQVTRVVTHEMAHADQWNAKNSDITVQYGGNNAALVNOTASDS 133  
 Db 63 GFGGGNNVSRKSTQD-MYTR-----AELEASAAANKEDFFARKRAENESRPEGLPPSQG 115  
 QY 134 SVMVRQVGFQGNATNQ 150  
 Db 116 G---KYVFGSGGPAPNQ 129

## RESULT 9

US-10-424-599-275468  
 ; Sequence 275468, Application US/10424599  
 ; Publication No. US20040031072A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: La Rosa Thomas J  
 ; APPLICANT: Kovalic David K  
 ; APPLICANT: Zhou Yihua  
 ; APPLICANT: Cao Yongwei  
 ; TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated With  
 ; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement  
 ; FILE REFERENCE: 38-21(53223)B  
 ; CURRENT APPLICATION NUMBER: US/10/424,599  
 ; CURRENT FILING DATE: 2003-04-28  
 ; NUMBER OF SEQ ID NOS: 285684  
 ; SEQ ID NO 275468  
 ; LENGTH: 486  
 ; TYPE: PRT  
 ; ORGANISM: Glycine max  
 ; FEATURE:  
 ; OTHER INFORMATION: Clone ID: PAT\_MRT3847\_90769C.1.pap  
 US-10-424-599-275468

Query Match 12.0%; Score 93; DB 12; Length 486;  
 Best Local Similarity 26.3%; Pred. No. 0.72;

Matches 36; Conservative 18; Mismatches 63; Indels 20; Gaps 6;  
 QY 14 VSGSALAGVVPQGGGNGHNGGNSGPDSTLSIYQVGSANAALQSDARKSETTITQS 73  
 Db 130 VRSSASKG-KPPLSAGNNNGGNDWDPD-----GFGSARGG-----ADLRNQSTGDYR 179  
 QY 74 GYNGADVGGADNYDQVTRVVTHEMAHADQWNAKNSDITVQYGGNNAALVNOTASDS 133  
 Db 180 GFGGGNNVSRKSTQD-MYTR-----AELEASAAANKEDFFARKRAENESRPEGLPPSQG 232  
 QY 134 SVMVRQVGFQGNATNQ 150  
 Db 233 G---KYVFGSGGPAPNQ 246

## RESULT 10

US-10-425-114-57763  
 ; Sequence 57763, Application US/10425114  
 ; Publication No. US20040034888A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Liu, Jingdong  
 ; APPLICANT: Zhou, Yihua  
 ; APPLICANT: Kovalic, David K.  
 ; APPLICANT: Screen, Steven E  
 ; APPLICANT: Tabaska, Jack E  
 ; APPLICANT: Cao, Yongwei  
 ; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With  
 ; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement  
 ; FILE REFERENCE: 38-21(53313)B  
 ; CURRENT APPLICATION NUMBER: US/10/425,114  
 ; CURRENT FILING DATE: 2003-04-28  
 ; NUMBER OF SEQ ID NOS: 73128  
 ; SEQ ID NO 57763  
 ; LENGTH: 507  
 ; TYPE: PRT  
 ; ORGANISM: Glycine max  
 ; FEATURE:  
 ; OTHER INFORMATION: Clone ID: UC-GMFLMINSOY049H04\_FLI.pap  
 US-10-425-114-57763

Query Match 12.0%; Score 93; DB 12; Length 507;  
 Best Local Similarity 26.3%; Pred. No. 0.76;  
 Matches 36; Conservative 18; Mismatches 63; Indels 20; Gaps 6;

QY 14 VSGSALAGVVPQGGGNGHNGGNSGPDSTLSIYQVGSANAALQSDARKSETTITQS 73  
 Db 151 VRSSASKG-KPPLSAGNNNGGNDWDPD-----GFGSARGG-----ADLRNQSTGDYR 200  
 QY 74 GYNGADVGGADNYDQVTRVVTHEMAHADQWNAKNSDITVQYGGNNAALVNOTASDS 133  
 Db 201 GFGGGNNVSRKSTQD-MYTR-----AELEASAAANKEDFFARKRAENESRPEGLPPSQG 253  
 QY 134 SVMVRQVGFQGNATNQ 150  
 Db 254 G---KYVFGSGGPAPNQ 267

## RESULT 11

US-09-880-748-1130  
 ; Sequence 1130, Application US/09880748  
 ; Publication No. US20030059937A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Ruben et al.  
 ; TITLE OF INVENTION: Antibodies that Immunospecifically Bind Blys  
 ; FILE REFERENCE: PF523  
 ; CURRENT APPLICATION NUMBER: US/09/880,748  
 ; CURRENT FILING DATE: 2001-06-15  
 ; PRIOR APPLICATION NUMBER: 60/212,210  
 ; PRIOR FILING DATE: 2000-06-15  
 ; PRIOR APPLICATION NUMBER: 60/240,816  
 ; PRIOR FILING DATE: 2000-10-17  
 ; PRIOR APPLICATION NUMBER: 60/276,248  
 ; PRIOR FILING DATE: 2001-03-16

```
; PRIOR APPLICATION NUMBER: 60/277,379
; PRIOR FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: 60/293,499
; PRIOR FILING DATE: 2001-05-25
; NUMBER OF SEQ ID NOS: 3239
; SOFTWARE: Patentin Ver. 2.0
; SEQ ID NO 1130
; LENGTH: 251
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-880-748-1130

Query Match      11.9%; Score 92.5; DB 10; Length 251;
Best Local Similarity 27.7%; Pred. No. 0.36; 51; Indels 21; Gaps 4;
Matches 33; Conservative 14; Mismatches 51; Indels 21; Gaps 4;

QY 18 ALAGVVPQWG-----GGGNHNGGSSGPDSTLSIYQYGSANAALALQSDARKSE- 67
Db 106 ATTGALDMWKGTLVTVSSGGGGGGGGGG-----GSAQAVLTQPSASGTPG 155

QY 68 TTITQSYGNGADVGQGAN-YDQLVTRVVTHEMAHADQNAKNSDITVQYGGNNAAL 125
Db 156 QRVTMSCGSSSSNIGSNVTNWYQQLPGAAPKLLIYRSDQRSSGVPDRFSGSKSGTASL 214

RESULT 12
US-10-293-418-1130
; Sequence 1130, Application US/10293418
; Publication No. US2003022396A1
; GENERAL INFORMATION:
; APPLICANT: Ruben et al.
; TITLE OF INVENTION: Antibodies that Immunospecifically Bind Blys
; FILE REFERENCE: PF523P2
; CURRENT APPLICATION NUMBER: US/10/293,418
; CURRENT FILING DATE: 2002-11-27
; PRIOR APPLICATION NUMBER: 60/331,469
; PRIOR FILING DATE: 2001-11-16
; PRIOR APPLICATION NUMBER: 60/340,817
; PRIOR FILING DATE: 2001-12-19
; PRIOR APPLICATION NUMBER: 09/880,748
; PRIOR FILING DATE: 2001-06-15
; PRIOR APPLICATION NUMBER: 60/293,499
; PRIOR FILING DATE: 2001-05-25
; PRIOR APPLICATION NUMBER: 60/277,379
; PRIOR FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: 60/276,248
; PRIOR FILING DATE: 2001-03-16
; PRIOR APPLICATION NUMBER: 60/240,816
; PRIOR FILING DATE: 2000-10-17
; PRIOR APPLICATION NUMBER: 60/212,210
; PRIOR FILING DATE: 2000-06-16
; NUMBER OF SEQ ID NOS: 3247
; SEQ ID NO 1130
; LENGTH: 251
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-293-418-1130

Query Match      11.9%; Score 92.5; DB 12; Length 251;
Best Local Similarity 27.7%; Pred. No. 0.36; 51; Indels 21; Gaps 4;
Matches 33; Conservative 14; Mismatches 51; Indels 21; Gaps 4;

QY 18 ALAGVVPQWG-----GGGNHNGGSSGPDSTLSIYQYGSANAALALQSDARKSE- 67
Db 106 ATTGALDMWKGTLVTVSSGGGGGGGGGG-----GSAQAVLTQPSASGTPG 155

QY 68 TTITQSYGNGADVGQGAN-YDQLVTRVVTHEMAHADQNAKNSDITVQYGGNNAAL 125
Db 156 QRVTMSCGSSSSNIGSNVTNWYQQLPGAAPKLLIYRSDQRSSGVPDRFSGSKSGTASL 214

RESULT 13
US-09-841-132-445
```

```
; Sequence 445, Application US/09841132
; Patent No. US20020061848A1
; GENERAL INFORMATION:
; APPLICANT: Bhatia, Ajay
; APPLICANT: Skeiky, Yasir A.W.
; APPLICANT: Probst, Peter
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR TREATMENT AND
; TITLE OF INVENTION: DIAGNOSIS OF CHLAMYDIAL INFECTION
; FILE REFERENCE: 210121.469C8
; CURRENT APPLICATION NUMBER: US/09/841,132
; CURRENT FILING DATE: 2001-04-23
; NUMBER OF SEQ ID NOS: 599
; SOFTWARE: FastSeq for Windows Version 3.0/4.0
; SEQ ID NO 445
; LENGTH: 1751
; TYPE: PRT
; ORGANISM: Chlamydia trachomatis serovar D
US-09-841-132-445

Query Match      11.7%; Score 91; DB 9; Length 1751;
Best Local Similarity 33.1%; Pred. No. 5.7;
Matches 40; Conservative 10; Mismatches 59; Indels 12; Gaps 5;

QY 15 SGSALAGVVPQWGGGNN--HNGGNSGPDSTLSIYQYGSANAALALQSDARKSETTIT- 71
Db 394 NADAWASSSPQSGGATTVNSGSDSGSDTSETVPATAKGG-GLYTDKNLSITNITG 452

QY 72 --QSGYNGADYGGQADNYDQLVTRVVTHEMAHADQNAKNSDITVGO-YGNNNAALVNQ 128
Db 453 IIEIANNKATDVGGGA-----YVKGTLTCENSHRLQFLKNSSDKGGGIYGEDNITLSNL 507

QY 129 T 129
Db 508 T 508

RESULT 14
US-09-841-132-594
; Sequence 594, Application US/09841132
; Patent No. US20020061848A1
; GENERAL INFORMATION:
; APPLICANT: Bhatia, Ajay
; APPLICANT: Skeiky, Yasir A.W.
; APPLICANT: Probst, Peter
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR TREATMENT AND
; TITLE OF INVENTION: DIAGNOSIS OF CHLAMYDIAL INFECTION
; FILE REFERENCE: 210121.469C8
; CURRENT APPLICATION NUMBER: US/09/841,132
; CURRENT FILING DATE: 2001-04-23
; NUMBER OF SEQ ID NOS: 599
; SOFTWARE: FastSeq for Windows Version 3.0/4.0
; SEQ ID NO 594
; LENGTH: 1751
; TYPE: PRT
; ORGANISM: C. Trachomatis D serovar
US-09-841-132-594

Query Match      11.7%; Score 91; DB 9; Length 1751;
Best Local Similarity 33.1%; Pred. No. 5.7;
Matches 40; Conservative 10; Mismatches 59; Indels 12; Gaps 5;

QY 15 SGSALAGVVPQWGGGNN--HNGGNSGPDSTLSIYQYGSANAALALQSDARKSETTIT- 71
Db 394 NADAWASSSPQSGGATTVNSGSDSGSDTSETVPATAKGG-GLYTDKNLSITNITG 452

QY 72 --QSGYNGADYGGQADNYDQLVTRVVTHEMAHADQNAKNSDITVGO-YGNNNAALVNQ 128
Db 453 IIEIANNKATDVGGGA-----YVKGTLTCENSHRLQFLKNSSDKGGGIYGEDNITLSNL 507

QY 129 T 129
Db 508 T 508
```

RESULT 15  
US-10-467-534-45  
; Sequence 45, Application US/10467534  
; Publication No. US20040131625A1  
; GENERAL INFORMATION:  
; APPLICANT: Berthet, Francois-Xavier Jacques  
; APPLICANT: Lobet, Yves  
; APPLICANT: Poolman, Jan  
; APPLICANT: Verlant, Vincent Georges Christian Louis  
; TITLE OF INVENTION: Vaccine Composition  
; FILE REFERENCE: B45261  
; CURRENT APPLICATION NUMBER: US/10/467,534  
; CURRENT FILING DATE: 2003-08-08  
; PRIOR APPLICATION NUMBER: PCT/EP02/01356  
; PRIOR FILING DATE: 2002-02-08  
; PRIOR APPLICATION NUMBER: GB 0103169.9  
; PRIOR FILING DATE: 2001-02-08  
; NUMBER OF SEQ ID NOS: 108  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 45  
; LENGTH: 1751  
; TYPE: PRT  
; ORGANISM: Chlamydia trachomatis  
US-10-467-534-45  
  
Query Match 11.7%; Score 91; DB 16; Length 1751;  
Best Local Similarity 33.1%; Pred. No. 5.7;  
Matches 40; Conservative 10; Mismatches 59; Indels 12; Gaps 5;  
  
QY 15 SGSALAGVVPQWGGGN--HNGGNSGSPDSTLSIYQGSANAALALQSDARKSETTIT- 71  
Db 394 NADAWASSSPQSGGATTVNSGSDSSGSDSTETVPATAGG-GLYTDKNLSITNITG 452  
QY 72 --QSGYGNGADVGGADNDYDQVTRVVTHEMAHADOWNAKNSDITVQY-YGGNNAALVNQ 128  
Db 453 IIEIANNKATDVGGGA-----YVKGTLTCENSHRLQFLKNSSDKQGGGIYGEDNITLSNL 507  
QY 129 T 129  
Db 508 T 508  
  
Search completed: August 2, 2004, 15:36:12  
Job time : 37.8 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: August 2, 2004, 14:48:33 ; Search time 167.9 Seconds  
(without alignments)  
877.809 Million cell updates/sec

Title: US-09-543-407-22  
Perfect score: 776  
Sequence: 1 MKLLKVAFAAIVSGSALA.....DSSVMRVQGVGNATANYQ 151

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 6019581 seqs, 976053577 residues  
Total number of hits satisfying chosen parameters: 6019581

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000  
Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Pending Patents AA Main: \*  
1: /cgn2\_6/prodata/2/paa/US06 COMB.pep: \*  
2: /cgn2\_6/prodata/2/paa/US06 COMB.pep: \*  
3: /cgn2\_6/prodata/2/paa/US080 COMB.pep: \*  
4: /cgn2\_6/prodata/2/paa/US081 COMB.pep: \*  
5: /cgn2\_6/prodata/2/paa/US082 COMB.pep: \*  
6: /cgn2\_6/prodata/2/paa/US083 COMB.pep: \*  
7: /cgn2\_6/prodata/2/paa/US084 COMB.pep: \*  
8: /cgn2\_6/prodata/2/paa/US085 COMB.pep: \*  
9: /cgn2\_6/prodata/2/paa/US086 COMB.pep: \*  
10: /cgn2\_6/prodata/2/paa/US087 COMB.pep: \*  
11: /cgn2\_6/prodata/2/paa/US088 COMB.pep: \*  
12: /cgn2\_6/prodata/2/paa/US089 COMB.pep: \*  
13: /cgn2\_6/prodata/2/paa/US090 COMB.pep: \*  
14: /cgn2\_6/prodata/2/paa/US091 COMB.pep: \*  
15: /cgn2\_6/prodata/2/paa/US092 COMB.pep: \*  
16: /cgn2\_6/prodata/2/paa/US093 COMB.pep: \*  
17: /cgn2\_6/prodata/2/paa/US094 COMB.pep: \*  
18: /cgn2\_6/prodata/2/paa/US095 COMB.pep: \*  
19: /cgn2\_6/prodata/2/paa/US096 COMB.pep: \*  
20: /cgn2\_6/prodata/2/paa/US097A COMB.pep: \*  
21: /cgn2\_6/prodata/2/paa/US097B COMB.pep: \*  
22: /cgn2\_6/prodata/2/paa/US098 COMB.pep: \*  
23: /cgn2\_6/prodata/2/paa/US099A COMB.pep: \*  
24: /cgn2\_6/prodata/2/paa/US099B COMB.pep: \*  
25: /cgn2\_6/prodata/2/paa/US100 COMB.pep: \*  
26: /cgn2\_6/prodata/2/paa/US101 COMB.pep: \*  
27: /cgn2\_6/prodata/2/paa/US102 COMB.pep: \*  
28: /cgn2\_6/prodata/2/paa/US103 COMB.pep: \*  
29: /cgn2\_6/prodata/2/paa/US104 COMB.pep: \*  
30: /cgn2\_6/prodata/2/paa/US106 COMB.pep: \*  
31: /cgn2\_6/prodata/2/paa/US107 COMB.pep: \*  
32: /cgn2\_6/prodata/2/paa/US107 COMB.pep: \*  
33: /cgn2\_6/prodata/2/paa/US60 COMB.pep: \*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match Length	ID	Description
------------	-------	--------------------	----	-------------

1	776	100.0	151	19	US-09-543-407-22	Sequence 22, Appl
2	712	91.8	151	19	US-09-543-407-28	Sequence 28, Appl
3	682	87.9	151	19	US-09-543-407-5	Sequence 5, Appl
4	677	87.2	151	6	US-08-233-642A-57	Sequence 57, Appl
5	657	84.7	151	19	US-09-543-407-30	Sequence 30, Appl
6	619	79.8	151	19	US-09-543-407-20	Sequence 20, Appl
7	609	77.5	151	19	US-09-543-407-14	Sequence 14, Appl
8	604	77.8	151	19	US-09-543-407-12	Sequence 12, Appl
9	602	77.6	151	19	US-09-543-407-14	Sequence 14, Appl
10	601	77.4	151	19	US-09-543-407-26	Sequence 26, Appl
11	600	77.3	151	19	US-09-543-407-18	Sequence 18, Appl
12	595	76.7	151	19	US-09-543-407-31	Sequence 31, Appl
13	567	73.1	151	19	US-09-543-407-16	Sequence 16, Appl
14	521	67.1	151	19	US-09-543-407-7	Sequence 7, Appl
15	518	66.8	151	13	US-08-978-878-4	Sequence 4, Appl
16	516	66.5	151	21	US-09-741-873B-4	Sequence 2, Appl
17	516	66.5	151	33	US-60-352-946-2	Sequence 2, Appl
18	516	66.5	151	33	US-60-444-371-2	Sequence 2, Appl
19	497	64.0	120	6	US-08-233-642A-55	Sequence 55, Appl
20	465	59.9	158	16	US-09-252-691C-5834	Sequence 5834, Ap
21	465	59.9	158	16	US-09-252-691C-5834	Sequence 5834, Ap
22	465	59.9	158	30	US-10-417-886-5834	Sequence 34, Appl
23	463	59.7	109	19	US-09-543-407-34	Sequence 2, Appl
24	440	56.7	131	13	US-08-978-878-2	Sequence 2, Appl
25	440	56.7	131	21	US-09-741-873B-2	Sequence 35, Appl
26	336	43.3	109	19	US-09-543-407-35	Sequence 37, Appl
27	256	33.0	68	19	US-09-543-407-37	Sequence 39, Appl
28	215	27.7	48	19	US-09-543-407-32	Sequence 32, Appl
29	159.5	20.6	70	19	US-09-543-407-32	Sequence 32, Appl
30	104.5	13.5	186	16	US-09-252-691C-5833	Sequence 5833, Ap
31	104.5	13.5	186	16	US-09-252-691C-5833	Sequence 5833, Ap
32	104.5	13.5	186	30	US-10-417-886-5833	Sequence 5833, Ap
33	97.5	12.6	520	1	PCT-US03-18256-21	Sequence 21, Appl
34	96	12.4	2308	33	US-60-161-932-688	Sequence 588, App
35	96	12.4	2309	20	US-09-614-150A-25488	Sequence 25488, A
36	96	12.4	2309	20	US-09-614-150A-25488	Sequence 25488, A
37	96	12.4	2309	33	US-60-191-637-25607	Sequence 25607, A
38	96	12.4	2309	33	US-60-191-681-20216	Sequence 20216, A
39	95.5	12.3	252	1	PCT-US03-13414-24	Sequence 24, Appl
40	95	12.2	445	29	US-10-369-493-20638	Sequence 20638, A
41	95	12.2	445	33	US-60-360-039-20638	Sequence 20638, A
42	95	12.2	597	1	PCT-US01-05992-146	Sequence 146, App
43	95	12.2	597	22	US-09-793-306-146	Sequence 146, App
44	94	12.1	271	30	US-10-437-963-147343	Sequence 147343,
45	93.5	12.0	145	21	US-09-739-449-8854	Sequence 8854, Ap

ALIGNMENTS

RESULT 1  
US-09-543-407-22  
; Sequence 22, Application US/09543407  
; GENERAL INFORMATION:  
; APPLICANT: White, Aaron P.  
; APPLICANT: Doran, James L.  
; APPLICANT: Collinson, S. Karen  
; APPLICANT: Kay, William W.  
; TITLE OF INVENTION: BACTERIAL FIMBRIAL SYSTEM FOR  
; PRESENTATION OF HETEROLOGOUS PEPTIDE SEQUENCES  
; FILE OF INVENTION: 920043.406  
; CURRENT APPLICATION NUMBER: US/09/543,407  
; CURRENT FILING DATE: 2000-04-05  
; NUMBER OF SEQ ID NOS: 59  
; SOFTWARE: FASTSEQ for Windows Version 4.0  
; SEQ ID NO 22  
; LENGTH: 151  
; TYPE: PRT  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: Recombinant Salmonella enteritidis 3b afga  
; OTHER INFORMATION: sequence containing the replacement fragment  
; OTHER INFORMATION: encoding PT3 from GP63 of Leishmania major.





```
QY 1 MKLLKVAAPAAIVVSGSALAGVVPQWGGGNGHNGGSSGPDSTLSIYQYGSANAALALQ 60
|
|
|
Db 1 MKLLKVAAPAAIVVSGSALAGVVPQWGGGNGHNGGSSGPDSTLSIYQYGSANAALALQ 60
|
|
|
QY 61 SDARKSETTITQSGYNGADVCGGADNYDQLVTRVVTHEMAHADQWNAKNSDITVQYGG 120
|
|
|
Db 61 SDARKSETTITQSGYNGADVCGGADNSTIELTQNGFRNATIDQWNAKNSDITVQYGG 120
|
|
|
QY 121 NNAALVNQTSASSVVMVRQVGFNNATANQY 151
|
|
|
Db 121 NNPALVNQTSASSVVMVRQVGFNNATANQY 151
|
|
|
RESULT 5
US-09-543-407-30
; Sequence 30, Application US/09543407
; GENERAL INFORMATION:
; APPLICANT: White, Aaron P.
; APPLICANT: Doran, James L.
; APPLICANT: Collinson, S. Karen
; APPLICANT: Kay, William W.
; TITLE OF INVENTION: BACTERIAL FIMBRIAL SYSTEM FOR
; PRESENTATION OF HETEROLOGOUS PEPTIDE SEQUENCES
; FILE REFERENCE: 920043.406
; CURRENT APPLICATION NUMBER: US/09/543,407
; CURRENT FILING DATE: 2000-04-05
; NUMBER OF SEQ ID NOS: 59
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 30
; LENGTH: 151
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Recombinant Salmonella enteritidis 3b afga
; OTHER INFORMATION: sequence containing the replacement fragment
; OTHER INFORMATION: encoding PT3 from GP63 of Leishmania major.
US-09-543-407-30

Query Match 84.7%; Score 657; DB 19; Length 151;
Best Local Similarity 81.9%; Pred. No. 7.4e-62;
Matches 136; Conservative 0; Mismatches 0; Indels 30; Gaps 2;

QY 1 MKLLKVAAPAAIVVSGSALAGVVPQWGGGNGHNGGSSGPDSTLSIYQYGSANAALALQ 60
|
|
|
Db 1 MKLLKVAAPAAIVVSGSALAGVVPQWGGGNGHNGGSSGPDSTLSIYQYGSANAALALQ 60
|
|
|
QY 61 SDARKSETTITQSGYNGADVCGGADN-----YDQLVTRVVTHEMAHADQ 105
|
|
|
Db 61 SDARKSETTITQSGYNGADVCGGADNSTIELTQNGFRNATIDQWNAKNSDITVQYGG 118
|
|
|
QY 106 WNAKNSDITVQYGGNNAALVNQTSASSVVMVRQVGFNNATANQY 151
|
|
|
Db 119 -----GNNALVNQTSASSVVMVRQVGFNNATANQY 151
|
|
|
RESULT 6
US-09-543-407-20
; Sequence 20, Application US/09543407
; GENERAL INFORMATION:
; APPLICANT: White, Aaron P.
; APPLICANT: Doran, James L.
; APPLICANT: Collinson, S. Karen
; APPLICANT: Kay, William W.
; TITLE OF INVENTION: BACTERIAL FIMBRIAL SYSTEM FOR
; PRESENTATION OF HETEROLOGOUS PEPTIDE SEQUENCES
; FILE REFERENCE: 920043.406
; CURRENT APPLICATION NUMBER: US/09/543,407
; CURRENT FILING DATE: 2000-04-05
; NUMBER OF SEQ ID NOS: 59
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 20
; LENGTH: 151
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Recombinant Salmonella enteritidis 3b afga
; OTHER INFORMATION: sequence containing the replacement fragment
; OTHER INFORMATION: encoding PT3 from GP63 of Leishmania major.
US-09-543-407-20

Query Match 78.5%; Score 609; DB 19; Length 151;
Best Local Similarity 73.6%; Pred. No. 1.1e-56;
Matches 128; Conservative 0; Mismatches 0; Indels 46; Gaps 2;

QY 1 MKLLKVAAPAAIVVSGSALAGVVPQWGGGNGHNGGSSGPDSTLSIYQYGSANAALALQ 60
|
|
|
Db 1 MKLLKVAAPAAIVVSGSALAGVVPQWGGGNGHNGGSSGPDSTLSIYQYGSANAALALQ 60
|
|
|
QY 61 SDARKSETTITQSGYNGADVCGGADN-----NYDQLVTRVVT 97
|
|
|
Db 61 SDARKSETTITQSGYNGADVCGGADNSTIELTQNGFRNATIDQWNAKNSDITVQYGG 120
|
|
|
QY 98 HEMAHADQWNAKNSDITVQYGGNNAALVNQTSASSVVMVRQVGFNNATANQY 151
|
|
|
Db 121 HEMAHA-----NOTASDSSVVMVRQVGFNNATANQY 151
|
|
|
RESULT 8
US-09-543-407-12
; Sequence 12, Application US/09543407
; GENERAL INFORMATION:
; APPLICANT: White, Aaron P.
; APPLICANT: Doran, James L.
```

```
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Recombinant Salmonella enteritidis 3b afga
; OTHER INFORMATION: sequence containing the replacement fragment
; OTHER INFORMATION: encoding PT3 from GP63 of Leishmania major.
US-09-543-407-20

Query Match 79.8%; Score 619; DB 19; Length 151;
Best Local Similarity 74.6%; Pred. No. 8.9e-58;
Matches 129; Conservative 0; Mismatches 0; Indels 44; Gaps 2;

QY 1 MKLLKVAAPAAIVVSGSALAGVVPQWGGGNGHNGGSSGPDSTLSIYQYGSANAALALQ 60
|
|
|
Db 1 MKLLKVAAPAAIVVSGSALAGVVPQWGGGNGHNGGSSGPDSTLSIYQYGSANAALALQ 60
|
|
|
QY 61 SDARKSETTITQSGYNGADVCGGADNYDQLVTRVVTHEMAHA-----103
|
|
|
Db 61 SDARK-----YDQLVTRVVTHEMAHAGGAGADNSTIELTQNGFR 98
|
|
|
QY 104 -----DQWNAKNSDITVQYGGNNAALVNQTSASSVVMVRQVGFNNATANQY 151
|
|
|
Db 99 NNAIDQWNAKNSDITVQYGGNNAALVNQTSASSVVMVRQVGFNNATANQY 151
|
|
|
RESULT 7
US-09-543-407-24
; Sequence 24, Application US/09543407
; GENERAL INFORMATION:
; APPLICANT: White, Aaron P.
; APPLICANT: Doran, James L.
; APPLICANT: Collinson, S. Karen
; APPLICANT: Kay, William W.
; TITLE OF INVENTION: BACTERIAL FIMBRIAL SYSTEM FOR
; PRESENTATION OF HETEROLOGOUS PEPTIDE SEQUENCES
; FILE REFERENCE: 920043.406
; CURRENT APPLICATION NUMBER: US/09/543,407
; CURRENT FILING DATE: 2000-04-05
; NUMBER OF SEQ ID NOS: 59
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 24
; LENGTH: 151
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Recombinant Salmonella enteritidis 3b afga
; OTHER INFORMATION: sequence containing the replacement fragment
; OTHER INFORMATION: encoding PT3 from GP63 of Leishmania major.
US-09-543-407-24

Query Match 78.5%; Score 609; DB 19; Length 151;
Best Local Similarity 73.6%; Pred. No. 1.1e-56;
Matches 128; Conservative 0; Mismatches 0; Indels 46; Gaps 2;

QY 1 MKLLKVAAPAAIVVSGSALAGVVPQWGGGNGHNGGSSGPDSTLSIYQYGSANAALALQ 60
|
|
|
Db 1 MKLLKVAAPAAIVVSGSALAGVVPQWGGGNGHNGGSSGPDSTLSIYQYGSANAALALQ 60
|
|
|
QY 61 SDARKSETTITQSGYNGADVCGGADN-----NYDQLVTRVVT 97
|
|
|
Db 61 SDARKSETTITQSGYNGADVCGGADNSTIELTQNGFRNATIDQWNAKNSDITVQYGG 120
|
|
|
QY 98 HEMAHADQWNAKNSDITVQYGGNNAALVNQTSASSVVMVRQVGFNNATANQY 151
|
|
|
Db 121 HEMAHA-----NOTASDSSVVMVRQVGFNNATANQY 151
|
|
|
RESULT 8
US-09-543-407-12
; Sequence 12, Application US/09543407
; GENERAL INFORMATION:
; APPLICANT: White, Aaron P.
; APPLICANT: Doran, James L.
```

APPLICANT: Collinson, S. Karen  
APPLICANT: Kay, William W.  
TITLE OF INVENTION: BACTERIAL FIMBRIAL SYSTEM FOR  
PRESENTATION OF HETEROLOGOUS PEPTIDE SEQUENCES  
FILE REFERENCE: 920043.406  
CURRENT APPLICATION NUMBER: US/09/543,407  
CURRENT FILING DATE: 2000-04-05  
NUMBER OF SEQ ID NOS: 59  
SOFTWARE: FastSEQ for Windows Version 4.0  
SEQ ID NO 12  
LENGTH: 151  
TYPE: PRT  
ORGANISM: Artificial Sequence  
FEATURE:  
OTHER INFORMATION: Recombinant Salmonella enteritidis 3b afga  
OTHER INFORMATION: sequence containing the replacement fragment  
OTHER INFORMATION: encoding PT3 from Gp63 of Leishmania major.  
US-09-543-407-12

Query Match 77.8%; Score 604; DB 19; Length 151;  
Best Local Similarity 80.8%; Pred. No. 3.6e-56;  
Matches 122; Conservative 6; Mismatches 23; Indels 0; Gaps 0;  
QY 1 MLLKVAFAAIVVSGSALAGVVPQGGGNGHNGSGSPDSTLSIYQYGSANAALALQ 60  
DB 1 MLLKVAFAAIVVSGSALAGVVPQGGGNGHNGSGSPDSTLSIYQYGSANAALALQ 60  
QY 61 SDARKSETTITQSGYNGADVGADNYDQVTRVVTHEMAHADQWNAKNSDITVGOYGG 120  
DB 61 SDARKSETTITQSGYNGADVGADNYDQVTRVVTHEMAHADQWNAKNSDITVGOYGG 120

QY 121 NNAALVNQTSADSSVMVQVGFNNATANY 151  
DB 121 NNAALVNDQVTRVVTHEMAHANNATANY 151

RESULT 9  
US-09-543-407-14  
Sequence 14, Application US/09543407  
GENERAL INFORMATION:  
APPLICANT: White, Aaron P.  
APPLICANT: Doran, James L.  
APPLICANT: Collinson, S. Karen  
TITLE OF INVENTION: BACTERIAL FIMBRIAL SYSTEM FOR  
PRESENTATION OF HETEROLOGOUS PEPTIDE SEQUENCES  
FILE REFERENCE: 920043.406  
CURRENT APPLICATION NUMBER: US/09/543,407  
CURRENT FILING DATE: 2000-04-05  
NUMBER OF SEQ ID NOS: 59  
SOFTWARE: FastSEQ for Windows Version 4.0  
SEQ ID NO 14  
LENGTH: 151  
TYPE: PRT  
ORGANISM: Artificial Sequence  
FEATURE:  
OTHER INFORMATION: Recombinant Salmonella enteritidis 3b afga  
OTHER INFORMATION: sequence containing the replacement fragment  
OTHER INFORMATION: encoding PT3 from Gp63 of Leishmania major.  
US-09-543-407-14

Query Match 77.6%; Score 602; DB 19; Length 151;  
Best Local Similarity 81.5%; Pred. No. 6e-56;  
Matches 123; Conservative 5; Mismatches 23; Indels 0; Gaps 0;  
QY 1 MLLKVAFAAIVVSGSALAGVVPQGGGNGHNGSGSPDSTLSIYQYGSANAALALQ 60  
DB 1 MLLKVAFAAIVVSGSALAGVVPQGGGNGHNGSGSPDSTLSIYQYGSANAALALQ 60  
QY 61 SDARKSETTITQSGYNGADVGADNYDQVTRVVTHEMAHADQWNAKNSDITVGOYGG 120  
DB 61 SDARKSETTITQSGYNGADVGADNYDQVTRVVTHEMAHADQWNAKNSDITVGOYGG 120

QY 121 NNAALVNQTSADSSVMVQVGFNNATANY 151  
DB 121 LVTRVVTHEMAHASVMVQVGFNNATANY 151

RESULT 10  
US-09-543-407-26  
Sequence 26, Application US/09543407  
GENERAL INFORMATION:  
APPLICANT: White, Aaron P.  
APPLICANT: Doran, James L.  
APPLICANT: Collinson, S. Karen  
APPLICANT: Kay, William W.  
TITLE OF INVENTION: BACTERIAL FIMBRIAL SYSTEM FOR  
PRESENTATION OF HETEROLOGOUS PEPTIDE SEQUENCES  
FILE REFERENCE: 920043.406  
CURRENT APPLICATION NUMBER: US/09/543,407  
CURRENT FILING DATE: 2000-04-05  
NUMBER OF SEQ ID NOS: 59  
SOFTWARE: FastSEQ for Windows Version 4.0  
SEQ ID NO 26  
LENGTH: 151  
TYPE: PRT  
ORGANISM: Artificial Sequence  
FEATURE:  
OTHER INFORMATION: Recombinant Salmonella enteritidis 3b afga  
OTHER INFORMATION: sequence containing the replacement fragment  
OTHER INFORMATION: encoding PT3 from Gp63 of Leishmania major.  
US-09-543-407-26

Query Match 77.4%; Score 601; DB 19; Length 151;  
Best Local Similarity 81.5%; Pred. No. 7.7e-56;  
Matches 123; Conservative 4; Mismatches 24; Indels 0; Gaps 0;  
QY 1 MLLKVAFAAIVVSGSALAGVVPQGGGNGHNGSGSPDSTLSIYQYGSANAALALQ 60  
DB 1 MLLKVAFAAIVVSGSALAGVVPQGGGNGHNGSGSPDSTLSIYQYGSANAALALQ 60  
QY 61 SDARKSETTITQSGYNGADVGADNYDQVTRVVTHEMAHADQWNAKNSDITVGOYGG 120  
DB 61 LVTRVVTHEMAHAGYNGADVGADNYDQVTRVVTHEMAHADQWNAKNSDITVGOYGG 120  
QY 121 NNAALVNQTSADSSVMVQVGFNNATANY 151  
DB 121 NNAALVNQTSADSSVMVQVGFNNATANY 151

RESULT 11  
US-09-543-407-18  
Sequence 18, Application US/09543407  
GENERAL INFORMATION:  
APPLICANT: White, Aaron P.  
APPLICANT: Doran, James L.  
APPLICANT: Collinson, S. Karen  
APPLICANT: Kay, William W.  
TITLE OF INVENTION: BACTERIAL FIMBRIAL SYSTEM FOR  
PRESENTATION OF HETEROLOGOUS PEPTIDE SEQUENCES  
FILE REFERENCE: 920043.406  
CURRENT APPLICATION NUMBER: US/09/543,407  
CURRENT FILING DATE: 2000-04-05  
NUMBER OF SEQ ID NOS: 59  
SOFTWARE: FastSEQ for Windows Version 4.0  
SEQ ID NO 18  
LENGTH: 151  
TYPE: PRT  
ORGANISM: Artificial Sequence  
FEATURE:  
OTHER INFORMATION: Recombinant Salmonella enteritidis 3b afga  
OTHER INFORMATION: sequence containing the replacement fragment  
OTHER INFORMATION: encoding PT3 from Gp63 of Leishmania major.  
US-09-543-407-18

Query Match 77.3%; Score 600; DB 19; Length 151;

Best Local Similarity 81.5%; Pred. No. 9.8e-56;  
Matches 123; Conservative 5; Mismatches 23; Indels 0; Gaps 0;  
QY 1 MKLLKVAAPAAIVVSGSALAGVVPQGGGNGHNGSGSPDSTLSIYQYGSANAALALQ 60  
DB 1 MKLLKVAAPAAIVVSGSALAGVVPQGGGNGHNGSGSPDSTLSIYQYGSANAALALQ 60  
QY 61 SDARKSETTITQSGYNGADVGQADNYDQVTRVVTHEMAHADQWNAKNSDITVGYGG 120  
DB 61 SDARKSETTITQSGYNGADVGQADNYDQVTRVVTHEMAHADQWNAKNSDITVGYGG 120  
QY 121 NNAALVNQTASDSSVMVRQVGFNNATANQY 151  
DB 121 NNAALVNQTASDSSVMVRQVGFNNATANQY 151

## RESULT 12

US-09-543-407-31  
; Sequence 31, Application US/09543407  
; GENERAL INFORMATION:  
; APPLICANT: White, Aaron P.  
; APPLICANT: Doran, James L.  
; APPLICANT: Collinson, S. Karen  
; APPLICANT: Kay, William W.  
; TITLE OF INVENTION: BACTERIAL FIMBRIAL SYSTEM FOR  
; FILE REFERENCE: 920043.406  
; CURRENT APPLICATION NUMBER: US/09/543,407  
; CURRENT FILING DATE: 2000-04-05  
; NUMBER OF SEQ ID NOS: 59  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 31  
; LENGTH: 131  
; TYPE: PRT  
; ORGANISM: Salmonella enteritidis  
US-09-543-407-31

Query Match 76.7%; Score 595; DB 19; Length 131;  
Best Local Similarity 89.3%; Pred. No. 2.8e-55;  
Matches 117; Conservative 2; Mismatches 12; Indels 0; Gaps 0;  
QY 21 GVVPQGGGNGHNGSGSPDSTLSIYQYGSANAALALQSDARKSETTITQSGYNGAD 80  
DB 1 GVVPQGGGNGHNGSGSPDSTLSIYQYGSANAALALQSDARKSETTITQSGYNGAD 60  
QY 81 VGGADNYDQVTRVVTHEMAHADQWNAKNSDITVGYGNNALVNQTASDSSVMVRQV 140  
DB 61 VGGADNSTIELTQNGFRNNATIDQWNAKNSDITVGYGNNALVNQTASDSSVMVRQV 120  
QY 141 GFGNNATANQY 151  
DB 121 GFGNNATANQY 131

## RESULT 13

US-09-543-407-16  
; Sequence 16, Application US/09543407  
; GENERAL INFORMATION:  
; APPLICANT: White, Aaron P.  
; APPLICANT: Doran, James L.  
; APPLICANT: Collinson, S. Karen  
; APPLICANT: Kay, William W.  
; TITLE OF INVENTION: BACTERIAL FIMBRIAL SYSTEM FOR  
; FILE REFERENCE: 920043.406  
; CURRENT APPLICATION NUMBER: US/09/543,407  
; CURRENT FILING DATE: 2000-04-05  
; NUMBER OF SEQ ID NOS: 59  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 16  
; LENGTH: 151  
; TYPE: PRT  
; ORGANISM: Artificial Sequence

FEATURE:  
; OTHER INFORMATION: Recombinant Salmonella enteritidis 3b afga  
; OTHER INFORMATION: sequence containing the replacement fragment  
; OTHER INFORMATION: encoding PT3 from GP63 of Leishmania major.  
US-09-543-407-16

Query Match 73.1%; Score 567; DB 19; Length 151;  
Best Local Similarity 80.8%; Pred. No. 3.4e-52;  
Matches 122; Conservative 5; Mismatches 24; Indels 0; Gaps 0;  
QY 1 MKLLKVAAPAAIVVSGSALAGVVPQGGGNGHNGSGSPDSTLSIYQYGSANAALALQ 60  
DB 1 MKLLKVAAPAAIVVSGSALAGVVDQVTRVVTHEMAHAGSPDSTLSIYQYGSANAALALQ 60  
QY 61 SDARKSETTITQSGYNGADVGQADNYDQVTRVVTHEMAHADQWNAKNSDITVGYGG 120  
DB 61 SDARKSETTITQSGYNGADVGQADNSTIELTQNGFRNNATIDQWNAKNSDITVGYGG 120  
QY 121 NNAALVNQTASDSSVMVRQVGFNNATANQY 151  
DB 121 NNAALVNQTASDSSVMVRQVGFNNATANQY 151

## RESULT 14

US-09-543-407-7  
; Sequence 7, Application US/09543407  
; GENERAL INFORMATION:  
; APPLICANT: White, Aaron P.  
; APPLICANT: Doran, James L.  
; APPLICANT: Collinson, S. Karen  
; APPLICANT: Kay, William W.  
; TITLE OF INVENTION: BACTERIAL FIMBRIAL SYSTEM FOR  
; FILE REFERENCE: 920043.406  
; CURRENT APPLICATION NUMBER: US/09/543,407  
; CURRENT FILING DATE: 2000-04-05  
; NUMBER OF SEQ ID NOS: 59  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 7  
; LENGTH: 151  
; TYPE: PRT  
; ORGANISM: Escherichia coli  
US-09-543-407-7

Query Match 67.1%; Score 521; DB 19; Length 151;  
Best Local Similarity 69.5%; Pred. No. 3e-47;  
Matches 105; Conservative 17; Mismatches 29; Indels 0; Gaps 0;  
QY 1 MKLLKVAAPAAIVVSGSALAGVVPQGGGNGHNGSGSPDSTLSIYQYGSANAALALQ 60  
DB 1 MKLLKVAAPAAIVVSGSALAGVVPQGGGNGHNGSGSPDSTLSIYQYGSANAALALQ 60  
QY 61 SDARKSETTITQSGYNGADVGQADNYDQVTRVVTHEMAHADQWNAKNSDITVGYGG 120  
DB 61 TDARNSDITITQGGGNGADVGQGSDDSSIDLTQGFNGSATLDQWNGKNSMTVKFGG 120  
QY 121 NNAALVNQTASDSSVMVRQVGFNNATANQY 151  
DB 121 GNGRAVDCTASNSSVMVTVQVGFNNATAHQY 151

## RESULT 15

US-08-978-878-4  
; Sequence 4, Application US/08978878  
; GENERAL INFORMATION:  
; APPLICANT: NORMARK, Staffan  
; APPLICANT: OLSEN, Arne  
; TITLE OF INVENTION: FIBRONECTIN BINDING PROTEIN AS WELL AS ITS PREPARATION  
; FILE REFERENCE: 012889-081  
; CURRENT APPLICATION NUMBER: US/08/978,878  
; CURRENT FILING DATE: 1997-11-26  
; EARLIER APPLICATION NUMBER: SE 8801723-1  
; EARLIER FILING DATE: 1988-05-06

```
; EARLIER APPLICATION NUMBER: US 07/347,189
; EARLIER FILING DATE: 1989-05-04
; EARLIER APPLICATION NUMBER: US 07/789,437
; EARLIER FILING DATE: 1991-11-06
; EARLIER APPLICATION NUMBER: US 07/970,846
; EARLIER FILING DATE: 1992-11-03
; EARLIER APPLICATION NUMBER: US 08/187,865
; EARLIER FILING DATE: 1994-01-28
; EARLIER APPLICATION NUMBER: US 08/319,519
; EARLIER FILING DATE: 1994-10-05
; EARLIER APPLICATION NUMBER: US 08/495,959
; EARLIER FILING DATE: 1995-06-28
; NUMBER OF SEQ ID NOS: 10
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 4
; LENGTH: 151
; TYPE: PRT
; ORGANISM: Escherichia coli
; US-08-978-878-4

Query Match      66.8%; Score 518; DB 13; Length 151;
Best Local Similarity 68.9%; Pred. No. 6.3e-47;
Matches 104; Conservative 18; Mismatches 29; Indels 0; Gaps 0;

QY      1 MKLLKVAAPAAIVVSGSALAGVVPQWGGGNGHNGGSSGPDSTLSIYOYGSANAALAIQ 60
Db      1 MKLLKVAALAAIVFSSAVAGVVPQYGGGNGHGGGNGSPNSELNIYQYGGNSALAIQ 60

QY      61 SPARKSETTITQSGYNGGADVGQGDADNYDQLVTRVVTHEMAHADQNAKNSDITVQYGG 120
Db      61 TDARNSDLTITQRCGGGADVGQGDSDSIDLTQRCFGNSATLDQWNGKNSMTVKQFEGG 120

QY      121 NNAAVYNOTASDSSVMVROVFGNNATANQY 151
Db      121 GNGAAVDQTASNSNVNTQVFGNNATAHQY 151
```

Search completed: August 2, 2004, 15:26:45  
Job time : 168.9 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: August 2, 2004, 14:49:38 ; Search time 17.8 Seconds  
(without alignments)  
888.146 Million cell updates/sec

Title: US-09-543-407-22

Perfect score: 776

Sequence: 1 MKLLKVAFAAIVSGSALA.....DSSVMVRQVGFNNATAHQY 151

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 601315 seqs, 104695340 residues

Total number of hits satisfying chosen parameters: 601315

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

Pending Patents AA New.\*  
1: /cgn2\_6/ptodata/2/paa/PCT\_NEW\_COMB.pep.\*  
2: /cgn2\_6/ptodata/2/paa/US06\_NEW\_COMB.pep.\*  
3: /cgn2\_6/ptodata/2/paa/US07\_NEW\_COMB.pep.\*  
4: /cgn2\_6/ptodata/2/paa/US08\_NEW\_COMB.pep.\*  
5: /cgn2\_6/ptodata/2/paa/US09\_NEW\_COMB.pep.\*  
6: /cgn2\_6/ptodata/2/paa/US10\_NEW\_COMB.pep.\*  
7: /cgn2\_6/ptodata/2/paa/US60\_NEW\_COMB.pep.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	518	66.8	151	US-09-741-873C-4	Sequence 4, Appli
2	440	56.7	131	US-09-741-873C-2	Sequence 2, Appli
3	97.5	12.6	520	US-10-479-638-21	Sequence 21, Appl
4	92	11.9	841	US-60-565-632-7906	Sequence 7906, Ap
5	92	11.9	841	US-60-579-062-7906	Sequence 7906, Ap
6	91	11.7	1751	US-10-498-327-115	Sequence 115, App
7	91	11.7	1751	US-10-872-155-445	Sequence 445, App
8	91	11.7	1751	US-10-872-155-594	Sequence 594, App
9	90	11.6	1010	PCT-US04-21492-402	Sequence 402, App
10	89.5	11.5	388	US-09-248-796A-17306	Sequence 17306, A
11	88.5	11.4	193	US-10-425-115-254240	Sequence 254240,
12	87	11.2	892	US-09-952-267B-5	Sequence 5, Appli
13	87	11.2	892	US-10-872-768-5	Sequence 5, Appli
14	87	11.2	892	US-10-872-768-5	Sequence 5, Appli
15	86.5	11.1	125	US-10-425-115-333419	Sequence 333419,
16	86.5	11.1	585	US-10-872-155-337	Sequence 337, App
17	86.5	11.1	1752	US-10-872-155-180	Sequence 180, App
18	86	11.1	201	US-10-425-115-309662	Sequence 309662,
19	86	11.1	244	US-10-854-439-84	Sequence 84, Appl
20	85	11.0	197	US-10-425-115-304391	Sequence 304391,
21	85	11.0	215	US-09-248-796A-20202	Sequence 20202, A
22	85	11.0	279	US-10-425-115-343835	Sequence 343835,
23	85	11.0	295	US-10-425-115-312468	Sequence 312468,
24	85	11.0	573	US-60-565-632-7907	Sequence 7907, Ap
25	85	11.0	573	US-60-579-062-7907	Sequence 7907, Ap
26	85	11.0	688	US-60-581-351-11388	Sequence 11388, A

27	84.5	10.9	282	6	US-10-501-282-2420	Sequence 2420, Ap
28	84.5	10.9	303	6	US-10-501-282-2422	Sequence 2422, Ap
29	84	10.8	256	6	US-10-425-115-301334	Sequence 301334,
30	84	10.8	619	1	PCT-US04-10229-38	Sequence 38, Appl
31	84	10.8	753	6	US-10-170-205E-35514	Sequence 35514, A
32	84	10.8	753	6	US-10-170-205E-35515	Sequence 35515, A
33	84	10.8	1627	6	US-10-170-205E-16659	Sequence 16659, A
34	84	10.8	1905	1	PCT-US04-09388-9	Sequence 9, Appli
35	83	10.7	585	1	PCT-US03-24982A-317	Sequence 317, App
36	82.5	10.6	246	6	US-10-854-439-511	Sequence 511, App
37	82.5	10.6	687	6	US-10-425-115-278239	Sequence 278239,
38	82	10.6	374	1	PCT-US04-11210-37	Sequence 285216,
39	81.5	10.5	508	6	US-10-425-115-285216	Sequence 285216,
40	81.5	10.5	532	6	US-10-425-115-285214	Sequence 18, Appl
41	81.5	10.5	669	6	US-10-489-425-18	Sequence 7905, Ap
42	81	10.4	412	7	US-60-565-632-7905	Sequence 7905, Ap
43	81	10.4	412	7	US-60-579-062-7905	Sequence 11, Appl
44	81	10.4	1358	6	US-10-778-804-11	Sequence 45603, A
45	80.5	10.4	234	6	US-10-767-701-45603	

#### ALIGNMENTS

RESULT 1  
US-09-741-873C-4  
; Sequence 4, Application US/09741873C  
; GENERAL INFORMATION:  
; APPLICANT: Normark, Staffan  
; TITLE OF INVENTION: Fibronectin Binding Protein As Well As Its Preparation  
; FILE REFERENCE: 012889-084  
; CURRENT APPLICATION NUMBER: US/09/741,873C  
; CURRENT FILING DATE: 2000-12-22  
; PRIOR APPLICATION NUMBER: SE 8801723-1  
; PRIOR FILING DATE: 1998-05-06  
; PRIOR APPLICATION NUMBER: US 08/978,878  
; PRIOR FILING DATE: 1997-11-26  
; PRIOR APPLICATION NUMBER: US 07/347,189  
; PRIOR FILING DATE: 1989-05-04  
; PRIOR APPLICATION NUMBER: US 07/789,437  
; PRIOR FILING DATE: 1991-11-06  
; PRIOR APPLICATION NUMBER: US 07/970,846  
; PRIOR FILING DATE: 1992-11-03  
; PRIOR APPLICATION NUMBER: US 08/187,865  
; PRIOR FILING DATE: 1994-01-28  
; PRIOR APPLICATION NUMBER: US 08/318,519  
; PRIOR FILING DATE: 1994-10-05  
; NUMBER OF SEQ ID NOS: 11  
; SOFTWARE: PatentIn version 3.0  
; SEQ ID NO 4  
; LENGTH: 151  
; TYPE: PRT  
; ORGANISM: Escherichia coli  
US-09-741-873C-4

Query Match 66.8%; Score 518; DB 5; Length 151;  
Best Local Similarity 68.9%; Pred No. 3.4e-37;  
Matches 104; Conservative 18; Mismatches 29; Indels 0; Gaps 0;

QY	1	MKLLKVAFAAIVSGSALA	GVVPMWGGGNNHNGGSGDSTLSIYQGSANALALQ	60
Db	1	MKLLKVAFAAIVSGSALA	GVVPMWGGGNNHNGGSGDSTLSIYQGSANALALQ	60
QY	61	SPARKSETITQSGYNGADYCGADNYDQLVTRVVTTHMAHADOWNAKNSDITVGYGG	120	
Db	61	TPARNSDLTITQHGNGGADYCGGSDSSIDLTQRGFGNSATLDQNGKNSMTVKQFGG	120	
QY	121	NNAALVNQTASDSSVMVRQVGFNNATAHQY	151	
Db	121	GNAAVDQTASNSVNVTVQVGFNNATAHQY	151	



APPLICANT: Wu, Wei  
APPLICANT: Zhang, Bei  
TITLE OF INVENTION: Methods for Genetic Control of Insect Infestations in Plants and  
FILE REFERENCE: 38-21 (53403) C  
CURRENT APPLICATION NUMBER: US/60/579,062  
CURRENT FILING DATE: 2004-06-11  
NUMBER OF SEQ ID NOS: 41445  
SOFTWARE: PatentIn version 3.2  
SEQ ID NO 7906  
LENGTH: 841  
TYPE: PRT  
ORGANISM: Diabrotica virgifera  
FEATURE:  
NAME/KEY: misc.feature  
LOCATION: (810)..(810)  
OTHER INFORMATION: xaa can be any naturally occurring amino acid  
US-60-579-062-7906

Query Match 11.7%; Score 92; DB 7; Length 841;  
Best Local Similarity 28.3%; Pred. No. 8.9;  
Matches 36; Conservative 16; Mismatches 59; Indels 16; Gaps 5;  
QY 30 GNHGGG--NSSGPDSTLSIYQGSANAALQSDARKSETTITQSG-----YQNGADV 81  
Db 657 GNHGGGSAENTGNASQNDAGQ--GSANAA-----NADNNANTDAQNGDQNGENGSAAEI 710  
QY 82 GCGADNYDQLVTRVVTHEMAHADQWNAKNSDITVQYGGNNAALVNOIASDSSVVRQVG 141  
Db 711 NGGADQNE--NGAAASGNDADQNNNNNSNDGNNQNGNSVEENSKEDSGNESQEN 768  
QY 142-FGNATA 148  
Db 769 KONNDIS 775

RESULT 6  
US-10-498-327-115  
Sequence 115, Application US/10498327  
GENERAL INFORMATION:  
APPLICANT: Grandi, Guido  
APPLICANT: Ratti, Giulio  
TITLE OF INVENTION: Immunisation Against Chlamydia Trachomatis  
FILE REFERENCE: 002441.00085  
CURRENT APPLICATION NUMBER: US/10/498,327  
CURRENT FILING DATE: 2004-06-10  
PRIOR APPLICATION NUMBER: PCT/IB02/05761  
PRIOR FILING DATE: 2002-12-12  
NUMBER OF SEQ ID NOS: 262  
SOFTWARE: PatentIn version 3.2  
SEQ ID NO 115  
LENGTH: 1751  
TYPE: PRT  
ORGANISM: Chlamydia trachomatis  
US-10-498-327-115

Query Match 11.7%; Score 91; DB 6; Length 1751;  
Best Local Similarity 33.1%; Pred. No. 25;  
Matches 40; Conservative 10; Mismatches 59; Indels 12; Gaps 5;  
QY 15 SGALAGVVPQWGGGN--HNGGNSGPDSTLSIYQGSANAALQSDARKSETTIT- 71  
Db 394 NADAWASSPQSGGATTVNSGSDSDTSETVPATKGG-GLYTDKNLSITNITG 452  
QY 72 --QSGYNGADVGGADNYDQLVTRVVTHEMAHADQWNAKNSDITVQY--YGGNNAALVNO 128  
Db 453 IIEIANNKATDVGGGA-----YVKGTLTCTENSHRLQFLKNSDKQGGGIYGEDNITLSNL 507  
QY 129 T 129  
Db 508 T 508

RESULT 7  
US-10-872-155-445  
Sequence 445, Application US/10872155  
GENERAL INFORMATION:  
APPLICANT: Bhatia, Ajay  
APPLICANT: Skeiky, Yasir A.W.  
APPLICANT: Probst, Peter  
TITLE OF INVENTION: COMPOUNDS AND METHODS FOR TREATMENT AND  
TITLE OF INVENTION: DIAGNOSIS OF CHLAMYDIAL INFECTION  
FILE REFERENCE: 210121.469C10  
CURRENT APPLICATION NUMBER: US/10/872,155  
CURRENT FILING DATE: 2004-06-18  
PRIOR APPLICATION NUMBER: 09/620,412  
PRIOR FILING DATE: 2000-07-20  
PRIOR APPLICATION NUMBER: 09/598,419  
PRIOR FILING DATE: 2000-06-20  
PRIOR APPLICATION NUMBER: 09/556,877  
PRIOR FILING DATE: 2000-04-19  
PRIOR APPLICATION NUMBER: 09/454,684  
PRIOR FILING DATE: 1999-12-03  
PRIOR APPLICATION NUMBER: 09/426,571  
PRIOR FILING DATE: 1999-10-22  
PRIOR APPLICATION NUMBER: 09/410,568  
PRIOR FILING DATE: 1999-10-01  
PRIOR APPLICATION NUMBER: 09/288,594  
PRIOR FILING DATE: 1999-04-08  
PRIOR APPLICATION NUMBER: 09/208,277  
PRIOR FILING DATE: 1998-12-08  
NUMBER OF SEQ ID NOS: 599  
SOFTWARE: FastSeq for Windows Version 3.0/4.0  
SEQ ID NO 445  
LENGTH: 1751  
TYPE: PRT  
ORGANISM: Chlamydia trachomatis serovar D  
US-10-872-155-445

Query Match 11.7%; Score 91; DB 6; Length 1751;  
Best Local Similarity 33.1%; Pred. No. 25;  
Matches 40; Conservative 10; Mismatches 59; Indels 12; Gaps 5;  
QY 15 SGALAGVVPQWGGGN--HNGGNSGPDSTLSIYQGSANAALQSDARKSETTIT- 71  
Db 394 NADAWASSPQSGGATTVNSGSDSDTSETVPATKGG-GLYTDKNLSITNITG 452  
QY 72 --QSGYNGADVGGADNYDQLVTRVVTHEMAHADQWNAKNSDITVQY--YGGNNAALVNO 128  
Db 453 IIEIANNKATDVGGGA-----YVKGTLTCTENSHRLQFLKNSDKQGGGIYGEDNITLSNL 507  
QY 129 T 129  
Db 508 T 508

RESULT 8  
US-10-872-155-594  
Sequence 594, Application US/10872155  
GENERAL INFORMATION:  
APPLICANT: Bhatia, Ajay  
APPLICANT: Skeiky, Yasir A.W.  
APPLICANT: Probst, Peter  
TITLE OF INVENTION: COMPOUNDS AND METHODS FOR TREATMENT AND  
TITLE OF INVENTION: DIAGNOSIS OF CHLAMYDIAL INFECTION  
FILE REFERENCE: 210121.469C10  
CURRENT APPLICATION NUMBER: US/10/872,155  
CURRENT FILING DATE: 2004-06-18  
PRIOR APPLICATION NUMBER: 09/620,412  
PRIOR FILING DATE: 2000-07-20  
PRIOR APPLICATION NUMBER: 09/598,419  
PRIOR FILING DATE: 2000-06-20  
PRIOR APPLICATION NUMBER: 09/556,877  
PRIOR FILING DATE: 2000-04-19  
PRIOR APPLICATION NUMBER: 09/454,684  
PRIOR FILING DATE: 1999-12-03

;; PRIOR APPLICATION NUMBER: 09/426,571  
;; PRIOR FILING DATE: 1999-10-22  
;; PRIOR APPLICATION NUMBER: 09/410,568  
;; PRIOR FILING DATE: 1999-10-01  
;; PRIOR APPLICATION NUMBER: 09/288,594  
;; PRIOR FILING DATE: 1999-04-08  
;; PRIOR APPLICATION NUMBER: 09/208,277  
;; PRIOR FILING DATE: 1998-12-08  
;; NUMBER OF SEQ ID NOS: 599  
;; SOFTWARE: FastSeq for Windows Version 3.0/4.0  
;; SEQ ID NO 594  
;; LENGTH: 1751  
;; TYPE: PRT  
;; ORGANISM: C. Trachomatis D serovar  
US-10-872-155-594

Query Match 11.7%; Score 91; DB 6; Length 1751;  
Best Local Similarity 33.1%; Pred. No. 25;  
Matches 40; Conservative 10; Mismatches 59; Indels 12; Gaps 5;  
QY 15 SGALAGVVPQWGGG--HNGGSGSPDSTLSIYQGSANAALALQSDARKSETTIT- 71  
Db 394 NADAWASSPQSGGATTVSNGSDSGSDSTETVPATAGG-GLYTDKLSITING 452  
QY 72 --OSGYGADVCGADNYDQVTRVTHMAHADOWNAKNSDITVQO-YGGNNAALVQ 128  
Db 453 IIEIANNKATDVGGGA-----YKGTLCENSHRLOFLKNSDKQGGGYGEDN-TLNL 507  
QY 129 T 129  
Db 508 T 508

RESULT 9  
PCT-US04-21492-402  
;; Sequence 402, Application PC/TUS0421492  
;; GENERAL INFORMATION:  
;; APPLICANT: Diversa Corporation  
;; APPLICANT: Steer, Brian  
;; APPLICANT: Callen, Walter  
;; APPLICANT: Healey, Shaun  
;; APPLICANT: Pulliam, Derrick  
;; TITLE OF INVENTION: GLUCANASES, NUCLEIC ACIDS ENCODING THEM AND METHODS FOR MAKING AN  
;; TITLE OF INVENTION: THEM  
;; FILE REFERENCE: 564462009540  
;; CURRENT APPLICATION NUMBER: PCT/US04/21492  
;; CURRENT FILING DATE: 2004-07-20  
;; PRIOR APPLICATION NUMBER: 60/484,725  
;; PRIOR FILING DATE: 2003-07-02  
;; NUMBER OF SEQ ID NOS: 518  
;; SOFTWARE: Patent In version 3.1  
;; SEQ ID NO 402  
;; LENGTH: 1010  
;; TYPE: PRT  
;; ORGANISM: Bacteria  
;; FEATURE:  
;; NAME/KEY: SIGNAL  
;; LOCATION: (1)...(30)  
;; FEATURE:  
;; NAME/KEY: DOMAIN  
;; LOCATION: (393)...(428)  
;; OTHER INFORMATION: Carbohydrate binding domain  
;; FEATURE:  
;; NAME/KEY: DOMAIN  
;; LOCATION: (39)...(300)  
;; OTHER INFORMATION: Cellulase (glycosyl hydrolase family 5)  
;; FEATURE:  
;; NAME/KEY: DOMAIN  
;; LOCATION: (610)...(959)  
;; OTHER INFORMATION: Glycosyl hydrolases family 6  
;; FEATURE:  
;; NAME/KEY: DOMAIN  
;; LOCATION: (493)...(521)

;; OTHER INFORMATION: Cellulose or protein binding domain  
PCT-US04-21492-402

Query Match 11.6%; Score 90; DB 1; Length 1010;  
Best Local Similarity 22.4%; Pred. No. 16;  
Matches 32; Conservative 27; Mismatches 56; Indels 28; Gaps 5;  
QY 12 IWSGSALAGVVPQW---GGNHNGGSGSPDSTLSIYQGSANAALALQSDARKSET 68  
Db 316 LTASGTLVKIVSNWGTIGNGSSSSSSSSSS-----SSSSSSSSSSSSSSSS 367  
QY 69 TITSGYNGA-----DVGQADNYDQVTRVTHMAHADOWN---NAKNSD-- 112  
Db 368 SSGSTGGGNCAGVNVYFNWTDWMSGAYNHANAGDMVYQNSLYRANWYTNVPGSDAS 427  
QY 113 -ITVQYGGNNAALVNOTASDSS 134  
Db 428 WTSLGACGGGSGTSSSSSSSSSS 450

RESULT 10  
US-09-248-796A-17306  
;; Sequence 17306, Application US/09248796A  
;; GENERAL INFORMATION:  
;; APPLICANT: Keith Weinstock et al  
;; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO CANDIDA ALBICANS  
;; TITLE OF INVENTION: FOR DIAGNOSTICS AND THERAPEUTICS  
;; FILE REFERENCE: 107196.132  
;; CURRENT APPLICATION NUMBER: US/09/248,796A  
;; CURRENT FILING DATE: 1999-02-12  
;; PRIOR APPLICATION NUMBER: US 60/074,725  
;; PRIOR FILING DATE: 1998-02-13  
;; PRIOR APPLICATION NUMBER: US 60/096,409  
;; PRIOR FILING DATE: 1998-08-13  
;; NUMBER OF SEQ ID NOS: 28208  
;; SEQ ID NO 17306  
;; LENGTH: 388  
;; TYPE: PRT  
;; ORGANISM: Candida albicans  
US-09-248-796A-17306

Query Match 11.5%; Score 89.5; DB 5; Length 388;  
Best Local Similarity 22.1%; Pred. No. 6;  
Matches 33; Conservative 17; Mismatches 44; Indels 55; Gaps 4;  
QY 30 GNHNGGSGSPDSTLSIYQGSANAALALQSDARKSETTITQSGYNGADVCGADNYD 89  
Db 132 GNQVGSNSYSPDT-----YGSAITGLVQEKTPAVTGIHSGIGAAA-YP 177  
QY 90 QLVTRVTHMAHADOWNAKNSDITVQYGGNNAALVNOTAS----- 131  
Db 178 ELT-----NAGNTGLAKGTAPASTATVYGESPSADYSKSGATGVVPATYLN 223  
QY 132 -----DSSVMVRQVGFNNATANOY 151  
Db 224 TSGAPTGLNTAGVVGAGGFGDNSNTSSY 252

RESULT 11  
US-10-425-115-254240  
;; Sequence 254240, Application US/10425115  
;; GENERAL INFORMATION:  
;; APPLICANT: La Rosa, Thomas J.  
;; APPLICANT: Kovalic, David K.  
;; APPLICANT: Zhou, Yihua  
;; APPLICANT: Cao, Yongwei  
;; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With  
;; TITLE OF INVENTION: Plants  
;; FILE REFERENCE: 38-21(53222)B  
;; CURRENT APPLICATION NUMBER: US/10/425,115  
;; CURRENT FILING DATE: 2003-04-28  
;; NUMBER OF SEQ ID NOS: 369326  
;; SEQ ID NO 254240



[illegible]

RESULT 15  
US-10-425-115-333419  
; Sequence 333419, Application US/10425115  
; GENERAL INFORMATION:  
; APPLICANT: La Rosa, Thomas J.  
; APPLICANT: Kovalic, David K.  
; APPLICANT: Zhou, Yihua  
; APPLICANT: Cao, Yongwei  
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With  
; TITLE OF INVENTION: Plants  
; FILE REFERENCE: 38-21(53222)B  
; CURRENT APPLICATION NUMBER: US/10/425,115  
; CURRENT FILING DATE: 2003-04-28  
; NUMBER OF SEQ ID NOS: 369326  
; SEQ ID NO 333419  
; LENGTH: 125  
; TYPE: PRT  
; ORGANISM: Zea mays  
; FEATURE:  
; OTHER INFORMATION: Clone ID: MRT4577\_67192C.1.pep  
US-10-425-115-333419

Query Match 11.1%; Score 86.5; DB 6; Length 125;  
Best Local Similarity 27.4%; Pred. No. 2.9;  
Matches 32; Conservative 14; Mismatches 44; Indels 27; Gaps 5;  
  
QY 27 GGGNHNHGGNSSGPDSTLSIYQGSANAALALQSDARKSETTI----- 70  
Db 1 GGYGAHYGG-----AAVAIQKGYRKDVYGETEDEDCEAFSPGRMQGGVQ 54  
  
QY 71 TQSGYGNADV--GQADNYDQIVTRVWTHEM--AHADQNAKNSDITVQYCGNNA 123  
Db 55 VQHYGGGVQVQHYGADNLGG--VRRYSYSEHGHGRRYSYKSTVQYNGGNA 110

Search completed: August 2, 2004, 15:29:54  
Job time : 18.8 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: August 2, 2004, 14:39:53 ; Search time 9.4 Seconds  
(without alignments)  
1545.204 Million cell updates/sec

Title: US-09-543-407-22

Perfect score: 776  
Sequence: 1 MKLLKVAFAAIVVSGSALA.....DSSVMVRQVFGNNATANCY 151

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 283366 seqs, 96191526 residues

Total number of hits satisfying chosen parameters: 283366

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

1: pir1.\*

2: pir2.\*

3: pir3.\*

4: pir4.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	682	87.9	151	2 JC6039	fimbrin protein ag
2	682	87.9	151	2 A10635	major curlin chain
3	521	67.1	151	2 S70788	curlin protein csg
4	493.5	64.4	152	2 D90806	curlin major subun
5	493.5	64.4	152	2 H85665	hypothetical prote
6	109	14.0	1748	2 S42136	cnjB protein - Tet
7	109	14.0	2174	2 E95925	hypothetical glyci
8	102	13.1	586	2 T26667	hypothetical prote
9	98	12.6	615	2 E70663	probable PPE prote
10	95	12.2	590	2 E70946	probable PPE prote
11	94	12.1	407	2 T21956	hypothetical prote
12	93.5	12.0	145	2 AD3143	conserved hypotet
13	93.5	12.0	145	2 H98144	hypothetical prote
14	93.5	12.0	645	2 P70825	probable PPE prote
15	93.5	12.0	1567	2 S11672	ice nucleation pro
16	92.5	11.9	151	2 S70787	curlin nucleate p
17	92.5	11.9	151	2 C90806	minor curlin subun
18	92.5	11.9	151	2 G85665	curlin minor chain
19	91.5	11.8	1034	2 JC2143	ice nucleation act
20	91.5	11.8	1258	2 JG0188	ice nucleation pro
21	91	11.7	1751	2 G71518	hypothetical prote
22	90.5	11.7	151	2 JC6040	fimbrin protein ag
23	90.5	11.7	151	2 A10635	nucleation compone
24	90.5	11.7	1053	2 E70987	probable PPE prote
25	89.5	11.5	1322	2 S07053	ice nucleation pro
26	89	11.5	575	2 S35327	protein kinase sgg
27	87	11.2	251	2 D96010	hypothetical expor
28	87	11.2	434	2 E70768	hypothetical glyci
29	87	11.2	573	2 C86266	F3F19.21 protein -

30 87 11.2 590 1 A45621 leishmanolysin (EC  
31 87 11.2 599 2 B42049 leishmanolysin (EC  
32 87 11.2 599 2 A44951 leishmanolysin (EC  
33 87 11.2 602 1 P10221 leishmanolysin (EC  
34 87 11.2 646 1 S19916 leishmanolysin (EC  
35 86.5 11.1 262 2 S00275 tail fiber protein  
36 86.5 11.1 487 2 C70830 DNA-binding protei  
37 86.5 11.1 1028 2 A56038 probable PPE prote  
38 86.5 11.1 1213 2 S16356 obo protein - frui  
39 86 11.1 582 2 F70675 probable PPE prote  
40 85.5 11.0 940 2 D89723 protein F39D8 1b  
41 85.5 11.0 945 2 T21998 hypothetical prote  
42 85.5 11.0 1210 2 A25347 ice nucleation pro  
43 85 11.0 354 2 B70663 probable PPE prote  
44 85 11.0 1052 2 AF2959 conserved hypotet  
45 85 11.0 1341 2 H98323 hypothetical prote

#### ALIGNMENTS

##### RESULT 1

JC6039  
fimbrin protein agfa precursor - Salmonella enteritidis  
C:Species: Salmonella enteritidis  
C:Date: 31-Dec-1996 #sequence revision 31-Dec-1996 #text\_change 08-Oct-1999  
C:Accession: JC6039; PC6015; A44898  
R:Collinson, S.K.; Clouthier, S.C.; Doran, J.L.; Baner, P.A.; Kay, W.W.  
J. Bacteriol. 178, 662-667, 1996  
A:Title: Salmonella enteritidis agfBAC operon encoding thin, aggregative fimbrinae.  
A:Reference number: JC6039; MUID:96146512; PMID:8550497  
A:Accession: JC6039  
A:Molecule type: DNA  
A:Residues: 1-151 <COL>  
A:Cross-references: GB:U43280; NID:G1184712; PIDN:AAC43599.1; PID:G1184714  
A:Accession: PC6015  
A:Molecule type: protein  
A:Residues: 21-52 <CO2>  
A:Experimental source: strain 27655-3b  
A:Note: the authors translated the codon ACG for residue 44 as Ile  
R:Collinson, S.K.; Emody, L.; Muller, K.H.; Trust, T.J.; Kay, W.W.  
J. Bacteriol. 173, 4773-4781, 1991  
A:Title: Purification and characterization of thin, aggregative fimbrinae from Salmonella  
A:Reference number: A44898; MUID:91310586; PMID:1677357  
A:Contents: 27655  
A:Accession: A44898  
A>Status: preliminary  
A:Molecule type: protein  
A:Residues: 21-33 <CO3>  
A:Note: sequence extracted from NCBI backbone (NCBIP:45936)  
C:Genetics:  
A:Gene: agfa  
C:Function:  
A:Description: major component of thin aggregative fimbrinae  
A:Note: fimbrinae bind to fibronectin, plasminogen, tissue plasminogen activator  
C:Keywords: fimbria  
F:1-20/Domain: signal sequence #status predicted <SIG>  
F:21-151/Product: fimbrin protein agfa #status experimental <MAT>

Query Match 87.9%; Score 682; DB 2; Length 151;  
Best Local Similarity 90.7%; Pred. No. 1.2e-50;  
Matches 137; Conservative 2; Mismatches 12; Indels 0; Gaps 0;

QY 1 MKLLKVAFAAIVVSGSALAGVVPQWGGGNGHNGGNSGPDSTLSITYQVGSANAALALQ 60  
DB 1 MKLLKVAFAAIVVSGSALAGVVPQWGGGNGHNGGNSGPDSTLSITYQVGSANAALALQ 60  
QY 61 SDARKSETTTTQSGYNGADYGGADNYDOLVTRVTHEMAHADQNAKNSDITVQYGG 120  
DB 61 SDARKSETTTTQSGYNGADYGGADNYDOLVTRVTHEMAHADQNAKNSDITVQYGG 120  
QY 121 NNAALVNTASDSSVMVRQVFGNNATANCY 151

Db 121 NNAALVNTASDSSVMVQVGFNNATANY 151

RESULT 2

AI0635

major curlin chain precursor [imported] - Salmonella enterica subsp. enterica serovar Typhimurium

C:Species: Salmonella enterica subsp. enterica serovar Typhi

C>Note: This species has also been called Salmonella typhi

C>Date: 09-Nov-2001 #sequence\_revision 09-Nov-2001 #text\_change 18-Nov-2002

C:Accession: AI0635

R:Parkhill, J.; Dougan, G.; James, K.D.; Thomson, N.R.; Pickard, D.; Wain, J.; Churcher, T.; Connor, P.; Cronin, A.; Davis, P.; Davies, R.M.; Dowd, L.; White, N.; Farrar, S.; Moulé, S.; O'Gaora, P.

Nature 413, 848-852, 2001

A:Authors: Parry, C.; Quail, M.; Rutherford, K.; Simmonds, M.; Skelton, J.; Stevens, K.

A>Title: Complete genome sequence of a multiple drug resistant Salmonella enterica serovar Typhimurium

A:Reference number: AB0502; MUID:21534947; PMID:11677608

A:Accession: AI0635

A>Status: preliminary

A:Molecule type: DNA

A:Residues: 1-151 <PAR>

A:Cross-references: GB:AL513382; PIDN:CAD08268.1; PID:gl6502315; GSPDB:GN00176

C:Genetics:

A:Gene: STY1181

Query Match 87.9%; Score 682; DB 2; Length 151;  
Best Local Similarity 90.7%; Pred. No. 1.2e-50;  
Matches 137; Conservative 2; Mismatches 12; Indels 0; Gaps 0;

QY 1 MLLKVAFAAIVVSGSALAGVVPQGGGNGHNGGNSGPDSTLSIYQGSANAALALQ 60

DB 1 MLLKVAFAAIVVSGSALAGVVPQGGGNGHNGGNSGPDSTLSIYQGSANAALALQ 60

QY 61 SDARKSETTITQSGYNGGADVGADNYDQVTRVVTHEMAHADOWNAKNSDITVQYGG 120

DB 61 SDARKSETTITQSGYNGGADVGADNYDQVTRVVTHEMAHADOWNAKNSDITVQYGG 120

QY 121 NNAALVNTASDSSVMVQVGFNNATANY 151

DB 121 NNAALVNTASDSSVMVQVGFNNATANY 151

RESULT 3

S70788

curlin protein csgA precursor - Escherichia coli (strain K-12)

N:Alternate names: csgA protein; major curlin protein

C:Species: Escherichia coli

C>Date: 12-Feb-1998 #sequence\_revision 20-Feb-1998 #text\_change 01-Mar-2002

C:Accession: S70788; G64846; S31202; S34560; S34559

R:Hammar, M.; Arngqvist, A.; Bian, Z.; Olsen, A.; Normark, S.

Mol. Microbiol. 18, 661-670, 1995

A>Title: Expression of two csg operons is required for production of fibronectin- and C

A:Reference number: S70788; MUID:96414468; PMID:8817489

A:Accession: S70788

A>Status: nucleic acid sequence not shown; translation not shown

A:Molecule type: DNA

A:Residues: 1-151 <HAM>

A:Cross-references: EMBL:X90754; NID:gl147558; PIDN:CAA62282.1; PID:gl147564

A:Experimental source: strain K12, substrain W3110

A>Note: The nucleotide sequence was submitted to the EMBL Data Library, August 1995

R:Blattner, F.R.; Plunkett III, G.; Bloch, C.A.; Perna, N.T.; Burland, V.; Riley, M.; Co

A.; Rose, D.J.; Mau, B.; Shaq, Y.

Science 277, 1453-1462, 1997

A>Title: The complete genome sequence of Escherichia coli K-12.

A:Reference number: A64720; MUID:97426617; PMID:9278503

A:Accession: G64846

A>Status: nucleic acid sequence not shown; translation not shown

A:Molecule type: DNA

A:Residues: 1-151 <BLAT>

A:Cross-references: GB:AE000205; GB:U00096; NID:gl1787265; PIDN:AAC74126.1; PID:gl1787279;

A:Experimental source: strain K-12, substrain MG1655

R:Olsen, A.; Arngqvist, A.; Hammar, M.; Sukupolvi, S.; Normark, S.

Mol. Microbiol. 7, 523-536, 1993

A>Title: The RpoS sigma factor relieves H-NS-mediated transcriptional repression of csgA,

A:Reference number: S31202; MUID:93211294; PMID:8459772

A:Accession: S31202

A:Molecule type: DNA

A:Residues: 1-6,'V',8-151 <OLS1>

A:Cross-references: EMBL:L04979

A:Accession: S34560

A:Molecule type: protein

A:Residues: 21-42;44-50 <OLS2>

R:Olsen, A.N.; Arngqvist, A.M.

submitted to the EMBL Data Library, October 1992

A:Reference number: S34559

A:Accession: S34559

C:Genetics:

A:Gene: csgA

A:Map position: 23.15

C:Function:

A>Description: major component of wild-type curli; interaction between CsgA and CsgB tri

A>Note: curli are thin, coiled fibers expressed on the surface of Escherichia coli that i

and H-kininogen; in the absence of CsgA, CsgB can self-assemble into polymers

F:1-20/Domain: signal sequence #status predicted <SIG>

F:21-151/Product: curlin #status experimental <MAT>

Query Match 67.1%; Score 521; DB 2; Length 151;  
Best Local Similarity 69.5%; Pred. No. 4.6e-37;  
Matches 105; Conservative 17; Mismatches 29; Indels 0; Gaps 0;

QY 1 MLLKVAFAAIVVSGSALAGVVPQGGGNGHNGGNSGPDSTLSIYQGSANAALALQ 60

DB 1 MLLKVAFAAIVVSGSALAGVVPQGGGNGHNGGNSGPDSTLSIYQGSANAALALQ 60

QY 61 SDARKSETTITQSGYNGGADVGADNYDQVTRVVTHEMAHADOWNAKNSDITVQYGG 120

DB 61 TDARNSDLTITQGGGNGADVGQSDSSIDLTRQFGNSATLDOWNKRNSEMTVQFGG 120

QY 121 NNAALVNTASDSSVMVQVGFNNATANY 151

DB 121 NNAALVNTASDSSVMVQVGFNNATANY 151

RESULT 4

D90806

curlin major subunit CsgA [imported] - Escherichia coli (strain O157:H7, substrain RIMD (

C:Species: Escherichia coli

C>Date: 18-Jul-2001 #sequence\_revision 18-Jul-2001 #text\_change 18-Jul-2001

C:Accession: D90806

R:Hayashi, T.; Makino, K.; Ohnishi, M.; Kurokawa, K.; Ishii, K.; Yokoyama, K.; Han, C.G.;

gasawara, N.; Yasunaga, T.; Kuhara, S.; Shiba, T.; Hattori, M.; Shinagawa, H.

DNA Res. 8, 11-22, 2001

A>Title: Complete genome sequence of enterohemorrhagic Escherichia coli O157:H7 and genom

A:Reference number: A99629; MUID:21156231; PMID:11258796

A:Accession: D90806

A>Status: preliminary

A:Molecule type: DNA

A:Residues: 1-152 <HAY>

A:Cross-references: GB:BA000007; PIDN:BA34843.1; PID:gl3360880; GSPDB:GN00154

A:Experimental source: strain O157:H7, substrain RIMD 0509952

C:Genetics:

A:Gene: ECs1420

Query Match 64.4%; Score 499.5; DB 2; Length 152;  
Best Local Similarity 67.8%; Pred. No. 3e-35;  
Matches 103; Conservative 18; Mismatches 30; Indels 1; Gaps 1;

QY 1 MLLKVAFAAIVVSGSALAGVVPQGGGNGHNGGNSGPDSTLSIYQGSANAALALQ 59

DB 1 MLLKVAFAAIVVSGSALAGVVPQGGGNGHNGGNSGPDSTLSIYQGSANAALALQ 60

QY 60 QSDARKSETTITQSGYNGGADVGADNYDQVTRVVTHEMAHADOWNAKNSDITVQYGG 119

[illegible]

## RESULT 8

T26667

hypothetical protein Y38E10A.q - Caenorhabditis elegans

C:Species: Caenorhabditis elegans

C:Date: 15-Oct-1999 #sequence\_revision 15-Oct-1999 #text\_change 15-Oct-1999

C:Accession: T26667

R:Wallis, J. Submitted to the EMBL Data Library, September 1999

A:Reference number: Z20252

A:Accession: T26667

A:Status: preliminary; translated from GB/EMBL/DDBJ

A:Molecule type: DNA

A:Residues: 1-586 &lt;WIL&gt;

A:Cross-references: EMBL:AL110484; NID:e1542205; PIDN:CAB54408.1; CESP:Y38E10A.q

A:Experimental source: Clone Y38E10A

C:Genetics:

A:Gene: CESP:Y38E10A.q

A:Introns: 105/2; 174/1; 248/1; 372/3; 436/3

Query Match

Best Local Similarity 13.1%; Score 102; DB 2; Length 586;

Matches 47; Conservative 14; Mismatches 66; Indels 36; Gaps 7;

QY 4 LKVAFAAIVVGSALAGVVPQWGGGNNH-----GGNSSGPDSTLSIYQGSANAALAL 59

DB 437 VSVGFGAQQFGSQPARPIAGGGGGGGGGGSGG-----YGAGGAGGAR 487

QY 60 QSDARKSETTITQSGYGCNAD----VGCADNYDQLVTRVT-----HEMAHADOWNAKN 110

DB 488 NSASNGS-----YGSANEVKSVGGAQQYGGSVFAPKPGTGCGYVAGSSARKSGE 539

QY 111 SDITVQYGGNNAALVNOTAS--DSSVMVRQVFGNNATANQY 151

DB 540 SGAGGGGAGGKAGGAKNSASYSSANEVKSVGFG---AQY 578

RESULT 9

E70663

probable PPE protein - Mycobacterium tuberculosis (strain H37RV)

C:Species: Mycobacterium tuberculosis

C:Date: 17-Jul-1998 #sequence\_revision 17-Jul-1998 #text\_change 15-Sep-2003

C:Accession: E70663

R:Coile, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon, S.

A:Connor, R.; Davies, R.; Devlin, K.; Feltwell, T.; Gentles, S.; Hamlin, N.; Holroyd, S.

Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S.

Nature 393, 537-544, 1998

A:Authors: Squares, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.

A:Title: Deciphering the biology of Mycobacterium tuberculosis from the complete genome

A:Reference number: A70500; MUID:98295987; PMID:9634230

A:Accession: E70663

A:Status: preliminary; nucleic acid sequence not shown; translation not shown

A:Molecule type: DNA

A:Residues: 1-615 &lt;COL&gt;

A:Cross-references: GB:Z83860; GB:AL123456; NID:g3261681; PIDN:CAB06165.1; PID:e290763;

A:Experimental source: strain H37RV

C:Genetics:

A:Gene: PPE

Query Match

Best Local Similarity 12.6%; Score 98; DB 2; Length 615;

Matches 31; Conservative 16; Mismatches 45; Indels 46; Gaps 5;

QY 15 SGALAGVVPQWGGGNNHGGSSGPDSTLSIYQGSANAALALQSDARKSETTITQSG 74

DB 365 STGNIG-----FNSGNNNIGFNSG-DGNIGFFNSGDN-----IG 401

QY 75 YGNGADVGGADNYDQLVTRVVTHEMAHADOWNAKNSDITVQYGGNNAALV---NOTAS 131

DB 402 FGNAGNINTGF-----WNAGLNTFGSAGNNGVGFDDGNSNG 441

QY 132 DSSVMVRQVFGNNATAN 149

DB 442 SFNVGFQNTGFGNSGACN 459

## RESULT 10

E70946

probable PPE protein - Mycobacterium tuberculosis (strain H37RV)

C:Species: Mycobacterium tuberculosis

C:Date: 17-Jul-1998 #sequence\_revision 17-Jul-1998 #text\_change 15-Sep-2003

C:Accession: E70946

R:Coile, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon, S.

A:Connor, R.; Davies, R.; Devlin, K.; Feltwell, T.; Gentles, S.; Hamlin, N.; Holroyd, S.

Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S.

Nature 393, 537-544, 1998

A:Authors: Squares, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.

A:Title: Deciphering the biology of Mycobacterium tuberculosis from the complete genome

A:Reference number: A70500; MUID:98295987; PMID:9634230

A:Accession: E70946

A:Status: preliminary; nucleic acid sequence not shown; translation not shown

A:Molecule type: DNA

A:Residues: 1-590 &lt;COL&gt;

A:Cross-references: GB:AL021646; GB:AL123456; NID:g3242278; PIDN:CAA16624.1; PID:e124875;

A:Experimental source: strain H37RV

C:Genetics:

A:Gene: PPE

Query Match

Best Local Similarity 12.2%; Score 95; DB 2; Length 590;

Matches 34; Conservative 14; Mismatches 50; Indels 26; Gaps 4;

QY 26 WGGGNNHGGSSGPDSTLSIYQGSANAALALQSDARKSETTITQSGYGGNADYQGA 85

DB 351 FNSGNNNIGFNSG-NNNVGFFNSGNNN-----FGFGNAGDINTGF 391

QY 86 DNYDQLVTRVVTHEMAHADOWNAKNSDITVQYGGNNAALVNOTASDSSVMVRQVFGN 145

DB 392 GNAGDTNTGFGNAGFFNMGIGNAGNEDMGVNGSGSFNVGVGN--AGNQS-----VGFGNA 444

QY 146 ATAN 149

DB 445 GTLN 448

## RESULT 11

T21956

hypothetical protein F38B7.3 - Caenorhabditis elegans

C:Species: Caenorhabditis elegans

C:Date: 15-Oct-1999 #sequence\_revision 15-Oct-1999 #text\_change 15-Oct-1999

C:Accession: T21956

R:Lennard, N.

submitted to the EMBL Data Library, June 1996

A:Reference number: Z19493

A:Accession: T21956

A:Status: preliminary; translated from GB/EMBL/DDBJ

A:Molecule type: DNA

A:Residues: 1-407 &lt;WIL&gt;

A:Cross-references: EMBL:Z74033; PIDN:CAA98477.1; GSPDB:GN00023; CESP:F38B7.3

A:Experimental source: clone F38B7

C:Genetics:

A:Gene: CESP:F38B7.3

A:Map position: 5

A:Introns: 12/1; 57/3; 124/2; 163/1; 330/3

Query Match

Best Local Similarity 12.1%; Score 94; DB 2; Length 407;

Matches 34; Conservative 19; Mismatches 43; Indels 36; Gaps 7;

QY 9 FAIVVSGSALA-GVVPQWGG-----GNNHGGSSGPDSTLSIYQGSANAALALQ 60

DB 250 FGSGLLAGSLGGLGSMWGGHHSGYGGGGYGGG-----YGMAG---GY 293

QY 61 SDARKSETTITQSGYGGNADYQGAADNYDQLVTRVVTHEMAHADOWNAKNSDITVQYGG 120

Db 294 SD---NDTVTNNYNGDNDNGNNHSS--SSNAITDAPGNAGDSQQSQSD----- 341

QY 121 NNAALVNQTASD 132

Db 342 NNYGNANQDSYD 353

RESULT 12

AD3143

conserved hypothetical protein Atu4768 [imported] - Agrobacterium tumefaciens (strain C58)

C:Species: Agrobacterium tumefaciens

C:Date: 11-Jan-2002 #sequence\_revision 11-Jan-2002 #text\_change 18-Nov-2002

C:Accession: AD3143

R:Wood, D.W.; Setubal, J.C.; Kaul, R.; Monks, D.; Chen, L.; Wood, G.E.; Chen, Y.; Woo, I.; Erage, G.; Gillet, W.; Grant, C.; Guenther, D.; Kutayavin, T.; Levy, R.; Li, M.; McClellan, Karp, P.; Romero, P.; Zhang, S.

Science 294, 2317-2323, 2001

A:Authors: Yoo, H.; Tao, Y.; Biddle, P.; Jung, M.; Krespan, W.; Perry, M.; Gordon-Kamm, S.; E.W.

A:Title: The Genome of the Natural Genetic Engineer Agrobacterium tumefaciens C58.

A:Reference number: AB2577; MUID:21608550; PMID:11743193

A:Accession: AD3143

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-145 <KUR>

A:Cross-references: GB:AE008689; PIDN:AA45562.1; PID:G17743277; GSPDB:GNC00187

A:Experimental source: strain C58 (Dupont)

C:Genetics:

A:Gene: Atu4768

A:Map position: linear chromosome

Query Match 12.0%; Score 93.5; DB 2; Length 145;

Best Local Similarity 21.8%; Pred. No. 0.48;

Matches 34; Conservative 25; Mismatches 64; Indels 33; Gaps 4;

QY 3 LLKVAFAAIVVSGSALAGVVPQGG-----GGNNGGSSGPDSTLSIYQGSANA 55

Db 1 MIRKSFASALVALVGLSAAAPAMANDVRIEYQGSNAGSAGGAEYGNRIYQNGYNR 60

QY 56 ALALQSDARKSETTITQSGYNGADVGO-GADNDYDQLVTRVVTTHMAHADQWNAKNSDIT 114

Db 61 IVG-----HQYGRNLNSAVQGEHDNYGSTTQ-----NGNENVAG 95

RESULT 13

H98144

hypothetical protein AGR\_L\_228 [imported] - Agrobacterium tumefaciens (strain C58, Cerec)

C:Species: Agrobacterium tumefaciens

C:Date: 22-Oct-2001 #sequence\_revision 22-Oct-2001 #text\_change 18-Nov-2002

C:Accession: H98144

R:Goodner, B.; Hinkle, G.; Gattung, S.; Miller, N.; Blanchard, M.; Quorollo, B.; Goldman, A.; Liu, F.; Wollam, C.; Allinger, M.; Doughty, D.; Scott, C.; Lappas, C.; Markelz, B.; Science 294, 2323-2328, 2001

A:Title: Genome Sequence of the Plant Pathogen and Biotechnology Agent Agrobacterium tumefaciens

A:Reference number: A97359; MUID:21608551; PMID:11743194

A:Accession: H98144

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-145 <KUR>

A:Cross-references: GB:AE007870; PIDN:AAK8682.1; PID:G15158413; GSPDB:GNC00170

C:Genetics:

A:Gene: AGR\_L\_228

A:Map position: linear chromosome

Query Match 12.0%; Score 93.5; DB 2; Length 145;

Best Local Similarity 21.8%; Pred. No. 0.48;

Matches 34; Conservative 25; Mismatches 64; Indels 33; Gaps 4;

QY 3 LLKVAFAAIVVSGSALAGVVPQGG-----GGNNGGSSGPDSTLSIYQGSANA 55

Db 1 MIRKSFASALVALVGLSAAAPAMANDVRIEYQGSNAGSAGGAEYGNRIYQNGYNR 60

QY 56 ALALQSDARKSETTITQSGYNGADVGO-GADNDYDQLVTRVVTTHMAHADQWNAKNSDIT 114

Db 61 IVG-----HQYGRNLNSAVQGEHDNYGSTTQ-----NGNENVAG 95

QY 115 VQYGGNNAALVNQTASDSSVMVRQVGFNNATNQ 150

Db 96 IQGFGSNHTTILTDGNGNIAAGVQVGRGGSANVSQ 131

RESULT 14

F70825

probable PPE protein - Mycobacterium tuberculosis (strain H37RV)

C:Species: Mycobacterium tuberculosis

C:Date: 17-Jul-1998 #sequence\_revision 17-Jul-1998 #text\_change 15-Sep-2003

C:Accession: F70825

R:Coile, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon, S.; Connor, R.; Davies, R.; Devlin, K.; Feltwell, T.; Gentles, S.; Hamlin, N.; Holroyd, S.; Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S. Nature 393, 537-544, 1998

A:Authors: Squares, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.

A:Title: Deciphering the biology of Mycobacterium tuberculosis from the complete genome

A:Reference number: A70500; MUID:98295987; PMID:9634230

A:Accession: F70825

A:Status: preliminary; nucleic acid sequence not shown; translation not shown

A:Molecule type: DNA

A:Residues: 1-645 <COL>

A:Cross-references: GB:AL021958; GB:AL123456; NID:G3261536; PIDN:CAA17522.1; PID:ei25329;

A:Experimental source: strain H37RV

C:Genetics:

A:Gene: PPE

Query Match 12.0%; Score 93.5; DB 2; Length 645;

Best Local Similarity 24.5%; Pred. No. 2.6;

Matches 39; Conservative 17; Mismatches 68; Indels 35; Gaps 6;

QY 15 SGSALAGVVPQ-----WGGGNGHNGGNSGPDSTLSIYQGSANAALQSDARKSETT 69

Db 237 SGNVGVGIPSSFNVGSGNIGNANVGGNSG-DNNTFGNFGNANIGNAGNMGSPAV 295

QY 70 ITQS-----GYNGADVCGADNDYDQLVTRVVTTHMAHADQWNAKNSDITVQVVG----- 119

Db 296 PTPGNGVNGIGNGGNGFGGNTGN-----ANIGLNVGDNVGGNGSGSYNFGF 345

QY 120 ---GNAALVNQTASDSSVMVRQVGFNNATNQ 149

Db 346 GNTGNNIGITGSGNIGFGGLNSGSGNIGFGNSGTGN 384

RESULT 15

S11672

ice nucleation protein - Xanthomonas campestris

C:Species: Xanthomonas campestris

C:Date: 21-Nov-1993 #sequence\_revision 10-Nov-1995 #text\_change 26-Aug-1999

C:Accession: S11672

R:Zhao, J.; Orser, C.S.

Mol. Gen. Genet. 223, 163-166, 1990

A:Title: Conserved repetition in the ice nucleation gene inX from Xanthomonas campestris

A:Reference number: S11672; MUID:91080859; PMID:2259339

A:Accession: S11672

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-1567 <ZHA>

A:Cross-references: EMBL:X52970; NID:G48531; PIDN:CAA37140.1; PID:G48532

C:Superfamily: ice nucleation protein

Query Match 12.0%; Score 93.5; DB 2; Length 1567;

Best Local Similarity 28.2%; Pred. No. 6.9;

Matches 42; Conservative 21; Mismatches 41; Indels 45; Gaps 10;

QY 34 GGGNSGPDSTLSIYQGSANAALQSDARKSETTITQSGYNGADVGGAD---NY-- 88

Db	383	GS	TG	TAG	AD	STL	-	I	A	G	S	T	Q	T	A	-----	G	G	E	S	S	L	T	-	A	G	Y	G	S	T	Q	T	A	R	Q	S	D	I	T	A	G	Y	G	S	432						
Qy	89	-----	D	Q	L	V	T	R	V	V	T	H	E	M	A	H	A	D	-----	Q	M	A	K	N	S	D	I	T	V	G	O	Y	G	-----	G	N	N	A	L	V	126										
Db	433	TG	TAG	AD	STL	I	A	G	S	T	Q	T	S	G	S	D	S	S	L	T	A	G	Y	G	S	T	Q	T	A	R	K	G	S	D	I	T	A	G	-	Y	G	S	T	A	G	S	D	S	L	I	491
Qy	127	-----	N	O	T	A	S	D	S	S	V	M	V	R	Q	V	G	F	G	N	N	A	T	A	N	Q	150																								
Db	492	AG	Y	G	S	T	Q	T	A	G	S	E	S	S	L	T	--	A	G	Y	G	S	T	Q	T	A	Q	518																							

Search completed: August 2, 2004, 14:56:24  
Job time : 10.4 secs



GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: August 2, 2004, 14:36:12 ; Search time 5.3 Seconds  
(without alignments)  
1483.508 Million cell updates/sec

Title: US-09-543-407-22

Perfect score: 776  
Sequence: 1 MKLLKVAAPAAIVSGSALA.....DSSVMVROVGFNNATANQY 151

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 141681 seqs, 52070155 residues

Total number of hits satisfying chosen parameters: 141681

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt\_42.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	682	87.9	151	1	CSGA_SALTY
2	521	67.1	151	1	CSGA_ECOLI
3	499.5	64.4	152	1	CSGA_ECO57
4	93.5	12.0	1567	1	ICEN_XANTC
5	92.5	11.9	151	1	CSGB_ECOLI
6	91.5	11.8	1034	1	ICEN_PANAN
7	91.5	11.8	1258	1	ICEN_ERWHE
8	91	11.7	1754	1	EMPE_CHLTR
9	90.5	11.7	151	1	CSGB_SALTY
10	90.5	11.7	151	1	CSGB_SALTY
11	89.5	11.5	1322	1	ICEA_PANAN
12	87	11.2	491	1	YK98_MYCTU
13	87	11.2	590	1	GP63_LEIDO
14	87	11.2	593	1	GP63_LEICH
15	87	11.2	602	1	GP63_LEIMA
16	87	11.2	646	1	GP63_LEIME
17	86.5	11.1	262	1	VG38_BP22
18	86.5	11.1	487	1	Y442_MYCTU
19	86.5	11.1	548	1	CEAK_ECOLI
20	86.5	11.0	1028	1	ICEN_PSRFL
21	85.5	11.0	1210	1	N189_SCHPO
22	85	11.0	1778	1	P033_MOUSE
23	84.5	10.9	495	1	YB1L_ECOLI
24	84.5	10.9	760	1	BUN2_DROME
25	84.5	10.9	1211	1	P033_RAT
26	83.5	10.8	497	1	YV96_YEAST
27	83	10.7	1140	1	P033_HUMAN
28	82	10.6	500	1	YF48_MYCTU
29	82	10.6	678	1	HWE1_HUMAN
30	81.5	10.5	392	1	YD70_MYCPN
31	81.5	10.5	737	1	FSH_DROME
32	81.5	10.5	2038	1	GRP1_ORVSA
33	81	10.4	165	1	GRP1_ORVSA

#### ALIGNMENTS

##### RESULT 1

ID	CSGA_SALTY	STANDARD	PRT	151 AA
AC	P55225			
DT	01-OCT-1996 (Rel. 34, Created)			
DT	01-OCT-1996 (Rel. 34, Last sequence update)			
DT	10-OCT-2003 (Rel. 42, Last annotation update)			
DE	Major curlin subunit precursor (Fimbrin SEF17).			
GN	CSGA OR AGFA OR STM1144 OR STY1181 OR T1776.			
OS	Salmonella typhimurium,			
OS	Salmonella typhi, and			
OS	Salmonella enteritidis.			
OC	Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;			
OC	Enterobacteriaceae; Salmonella.			
OX	NCBI_TaxID=602, 601, 592;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RC	SPECIES=S.typhimurium; STRAIN=SR-11;			
RX	MEDLINE=98117058; PubMed=9457880;			
RA	Romling U., Bian Z., Hammar M., Sierralta W.D., Normark S.;			
RT	"Curli fibers are highly conserved between Salmonella typhimurium and			
RT	Escherichia coli with respect to operon structure and regulation.";			
RL	J. Bacteriol. 180:722-731(1998).			
RN	[2]			
RP	SEQUENCE FROM N.A.			
RC	SPECIES=S.typhimurium; STRAIN=LT2 / SGSC1412 / ATCC 700720;			
RX	MEDLINE=21534948; PubMed=11677609;			
RA	McClelland M., Sanderson K.E., Spieth J., Clifton S.W., Latreille P.,			
RA	Courtney L., Porwollik S., Ali J., Dante M., Du F., Hou S., Layman D.,			
RA	Leonard S., Nguyen C., Scott K., Holmes A., Grewal N., Mulvaney E.,			
RA	Ryan E., Sun H., Florea L., Miller W., Stoneking T., Nhan M.,			
RA	Waterston R., Wilson R.K.;			
RT	"Complete genome sequence of Salmonella enterica serovar Typhimurium			
RT	LT2.";			
RL	Nature 413:852-856 (2001).			
RN	[3]			
RP	SEQUENCE FROM N.A.			
RC	SPECIES=S.typhi; STRAIN=CT18;			
RX	MEDLINE=21534947; PubMed=11677608;			
RA	Parkhill J., Dougan G., James K.D., Thomson N.R., Pickard D., Wain J.,			
RA	Churcher C., Mungall K.L., Bentley S.D., Holden M.T.G., Sebaihia M.,			
RA	Baker S., Basham D., Brooks K., Chillingworth T., Connor P.,			
RA	Cronin A., Davis P., Davies R.M., Dowd L., White N., Farrar J.,			
RA	Feltwell T., Hamlin N., Haque A., Hien T.T., Holroyd S., Jagsis K.,			
RA	Krogh A., Larsen T.S., Leather S., Moule S., O'Gaora P., Parry C.,			
RA	Quail M.A., Rutherford K., Simmonds M., Skelton J., Stevens K.,			
RA	Whitehead S., Barrett B.G.;			
RT	"Complete genome sequence of a multiple drug resistant Salmonella			
RT	enterica serovar Typhi CT18.";			
RL	Nature 413:848-852 (2001).			
RN	[4]			
RP	SEQUENCE FROM N.A.			
RC	SPECIES=S.typhi; STRAIN=Ty2 / ATCC 700931;			
RX	MEDLINE=22511357; PubMed=12644504;			
RA	Deng W., Liou S.-R., Plunkett G. III, Mayhew G.F., Rose D.J.,			
RA	Burland V., Kodoyanni V., Schwartz D.C., Blattner F.R.;			

34	80.5	10.4	874	1	ALE6_AZOVI
35	80.5	10.4	1196	1	ICEV_PSEX
36	80	10.3	204	1	CORA_MEDSA
37	80	10.3	362	1	ALEI_STACP
38	80	10.3	429	1	DR48_YEAST
39	80	10.3	443	1	Y878_MYCTU
40	80	10.3	543	1	YP91_MYCTU
41	79.5	10.2	396	1	PER_DROPV
42	78.5	10.1	346	1	RO21_XENLA
43	78.5	10.1	1148	1	ICEK_PSEX
44	78.5	10.1	1317	1	N145_YEAST
45	78	10.1	362	1	P35_MYCPE

Q9zfh0	azotobacter
O33479	pseudomonas
Q07202	staphylococ
O05156	staphylococ
P18899	saccharomyc
Q10540	mycobacteri
Q50630	mycobacteri
P91698	drosophila
P51989	xenopus lae
O30611	pseudomonas
P49687	saccharomyc
Q50367	mycoplasma

RT "Comparative genomics of Salmonella enterica serovar Typhi strains Ty2  
 RL and Ctl8.";  
 RN J. Bacteriol. 185:2330-2337(2003).  
 [5]  
 RP SEQUENCE FROM N.A.  
 RC SPECIES=S.enteritidis; STRAIN=27655-3B;  
 RX MEDLINE=96146512; PubMed=8550497;  
 RA Collinson S.K., Clouthier S.C., Doran J.L., Bansen P.A., Kay W.W.;  
 RT "Salmonella enteritidis afBAC operon encoding thin, aggregative  
 fimbriae.";  
 RL J. Bacteriol. 178:662-667(1996).  
 [6]  
 RP SEQUENCE OF 21-151 FROM N.A.  
 RC SPECIES=S.enteritidis; STRAIN=27655-3B;  
 RX MEDLINE=91310586; PubMed=1677357;  
 RA Doran J.L., Collinson S.K., Burian J., Sarlos G., Todd B.C.D.,  
 PA Munro C.K., Kay C.W., Bansen P.A., Peterkin P.I., Kay W.W.;  
 RT "DNA-based diagnostic tests for Salmonella species targeting agfA,  
 RT the structural gene for thin, aggregative fimbriae.";  
 RL J. Clin. Microbiol. 31:2263-2273(1993).  
 [7]  
 RP SEQUENCE OF 21-33.  
 RC SPECIES=S.enteritidis; STRAIN=27655-3B;  
 RX MEDLINE=91310586; PubMed=1677357;  
 RA Collinson S.K., Emeody L., Mueller K.M., Trust T.J., Kay W.W.;  
 RT "Purification and characterization of thin, aggregative fimbriae from  
 RT Salmonella enteritidis.";  
 RL J. Bacteriol. 173:4773-4781(1991).  
 CC -!- FUNCTION: CURLIN IS THE STRUCTURAL SUBUNIT OF THE CURLI. CURLI ARE  
 CC COILED SURFACE STRUCTURES THAT ASSEMBLE PREFERENTIALLY AT GROWTH  
 CC TEMPERATURES BELOW 37 DEGREES CELSIUS. CURLI CAN BIND TO  
 CC FIBRONECTIN.  
 CC -!- SIMILARITY: BELONGS TO THE CSGA/CSGB FAMILY.  
 CC  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@sib-sib.ch](mailto:license@sib-sib.ch)).  
 CC  
 DR EMBL; AJ002301; CRA05317.1; -;  
 DR EMBL; AE008749; AL20074.1; -;  
 DR EMBL; AL627269; CAD08268.1; -;  
 DR EMBL; AL016840; AAC069399.1; -;  
 DR EMBL; U43280; AAC43599.1; -;  
 DR PIR; JC6039; JC6039.  
 DR StyGene; SG10608; csGA.  
 DR Fimbrin; Signal; Complete proteome.  
 FT SIGNAL 1 20  
 FT CHAIN 21 151 MAJOR CURLIN SUBUNIT.  
 FT CONFLICT 134 151 SVMVRQVGFNNATANYQ -> DSYTQVAS (IN  
 FT REF. 6).  
 FT SEQUENCE 151 AA; 15305 MW; B7DAC0D16B621359 CRC64;  
 SQ  
 Query Match 87.9%; Score 682; DB 1; Length 151;  
 Best Local Similarity 90.7%; Pred No. 2.2e-50;  
 Matches 137; Conservative 2; Mismatches 12; Indels 0; Gaps 0;  
 QY 1 MKLLKVAFAAIVVSGSALAGVVPQWGGNNGGNSGPDSTLSIYQYGSANAALALQ 60  
 DB 1 MKLLKVAFAAIVVSGSALAGVVPQWGGNNGGNSGPDSTLSIYQYGSANAALALQ 60  
 QY 61 SPARKSETTITGSGYNGADVGCGADNYDQLVTRVVTHEMADOWNAKNSDITVGYGG 120  
 DB 61 SPARKSETTITGSGYNGADVGCGADNSTIETQNGFRNNATIDOWNAKNSDITVGYGG 120  
 QY 121 NNAALVNTASDSSVMVRQVGFNNATANYQ 151  
 DB 121 NNAALVNTASDSSVMVRQVGFNNATANYQ 151

RESULT 2  
 CSGA\_ECOLI  
 ID CSGA\_ECOLI STANDARD; PRT; 151 AA.  
 AC P28307;  
 DT 01-DEC-1992 (Rel. 24, Created)  
 DT 01-OCT-1996 (Rel. 34, Last sequence update)  
 DT 28-FEB-2003 (Rel. 41, Last annotation update)  
 DE Major curlin subunit precursor.  
 GN CSGA OR B1042.  
 OS Escherichia coli.  
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;  
 OC Enterobacteriaceae; Escherichia.  
 OX NCBI TaxID=562;  
 [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=K12 / W3110;  
 RX MEDLINE=93211294; PubMed=8459772;  
 RA Olsen A., Arngvist A.;  
 RT "The RpoS sigma factor relieves H-NS-mediated transcriptional  
 RT repression of csGA, the subunit gene of fibronectin-binding curli in  
 RT Escherichia coli.";  
 RL Mol. Microbiol. 7:523-536(1993).  
 [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=K12 / MC4100;  
 RX MEDLINE=96414468; PubMed=8817489;  
 RA Hammar M., Arngvist A., Bian Z., Olsen A., Normark S.;  
 RT "Expression of two csG operons is required for production of  
 RT fibronectin- and congo red-binding curli polymers in Escherichia coli  
 RT K-12.";  
 RL Mol. Microbiol. 18:661-670(1995).  
 [3]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=K12 / MG1655;  
 RX MEDLINE=97426617; PubMed=9278503;  
 RA Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,  
 RA Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,  
 RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,  
 RA Mau B., Shao Y.;  
 RT "The complete genome sequence of Escherichia coli K-12.";  
 RL Science 277:1232-1244(1997).  
 [4]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=K12;  
 RX MEDLINE=97061202; PubMed=8905232;  
 RA Oshima T., Aiba H., Baba T., Fujita K., Hayashi K., Honjo A.,  
 RA Ikemoto K., Inada T., Itoh T., Kajihara M., Kanai K., Kashimoto K.,  
 RA Kimura S., Kitagawa M., Makino K., Masuda S., Miki T., Mizobuchi K.,  
 RA Mori H., Motomura K., Nakamura Y., Nishimoto H., Nishio Y., Saito N.,  
 RA Sampei G., Seki Y., Tagami H., Takemoto K., Wada C., Yamamoto Y.,  
 RA Yano M., Horiuchi T.;  
 RT "A 718-kb DNA sequence of the Escherichia coli K-12 genome  
 RT corresponding to the 12.7-28.0 min region on the linkage map.";  
 RL DNA Res. 3:137-155(1996).  
 [5]  
 RP SEQUENCE OF 21-40.  
 RC STRAIN=K12 / YMBL;  
 RX MEDLINE=93023873; PubMed=1357528;  
 RA Arngvist A., Olsen A., Pfeifer J., Russell D.G., Normark S.;  
 RT "The Crl protein activates cryptic genes for curli formation and  
 RT fibronectin binding in Escherichia coli HB101.";  
 RL Mol. Microbiol. 6:2443-2452(1992).  
 [6]  
 RP SEQUENCE OF 21-31.  
 RC MEDLINE=91310586; PubMed=1677357;  
 RA Collinson S.K., Emeody L., Trust T.J., Kay W.W.;  
 RT "Purification and characterization of thin, aggregative fimbriae from  
 RT Salmonella enteritidis.";  
 RL J. Bacteriol. 173:4773-4781(1991).  
 CC -!- FUNCTION: CURLIN IS THE STRUCTURAL SUBUNIT OF THE CURLI. CURLI ARE  
 CC COILED SURFACE STRUCTURES THAT ASSEMBLE PREFERENTIALLY AT GROWTH  
 CC TEMPERATURES BELOW 37 DEGREES CELSIUS. CURLI CAN BIND TO  
 CC FIBRONECTIN.

```

CC -!- SIMILARITY: BELONGS TO THE CSGA/CSGB FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; L04979; AAA23616.1; -
CC EMBL; X90754; CAA62282.1; -
CC EMBL; AE000205; AAC7426.1; -
CC EMBL; D90741; BAA3832.1; -
CC EMBL; D90742; BAA35840.1; -
CC PIR; S70788; S70788.
CC EcoGene; EG11489; CSGA.
CC Fimbria; Signal; Complete proteome.
CC SIGNAL 1 20 MAJOR CURLIN SUBUNIT.
CC CHAIN 21 151
CC CONFLICT 7 7 A -> E (IN REF. 1).
CC SEQUENCE 151 AA; 15049 MW; C003470D20BD395F CRC64;

Query Match 67.1%; Score 521; DB 1; Length 151;
Best Local Similarity 69.5%; Pred. No. 6e-37;
Matches 105; Conservative 17; Mismatches 29; Indels 0; Gaps 0;

Qy 1 MKLLKVAAPAAIVVSGSALAGVVPQWGGGNGHNGGNSGPDSTLSIYQYGSAALALQ 60
Db 1 MKLLKVAALAAIVFSGSALAGVVPQYGGGNGHNGGNSGPNSELNIYQYGGNSALALQ 60
Qy 61 SDARKSETTITQSGYGNAGDVGGADNYDQLVTRVVTTHMAHADOWNAKNSDITVQYGG 120
Db 61 TDARNSDLITQHGCGNGADVGGSDSSIDLQTFGFGNSATLDQWNGKNSMTVKQFGG 120

Qy 121 NNAALVNOTASDSVMVQVGFNNATANOY 151
Db 121 GNGAAVDQTASNSVNVTVQVGFNNATAHOY 151

RESULT 3
CSGA_ECO57 STANDARD; PRT; 152 AA.
AC Q93U24;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DE Major curlin subunit precursor.
GN CSGA OR 21676 OR ECS1420.
OS Escherichia coli O157:H7.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Escherichia.
OX NCBI_TaxID=83334;
RN SEQUENCE FROM N.A.
RP STRAIN=O157:H7 / ATCC 43895;
RC MEDLINE=21074935; PubMed=11319125;
RA Ulrich G.A., Keen J.E., Elder R.O.;
RE "Mutations in the csgeD promoter associated with variations in curli
RT expression in certain strains of Escherichia coli O157:H7.";
RL Appl. Environ. Microbiol. 67:2367-2370 (2001).
RN SEQUENCE FROM N.A.
RP STRAIN=O157:H7 / EDL933 / ATCC 700927;
RC MEDLINE=21074935; PubMed=11206551;
RA Perna N.T., Plunkett G. III, Burland V., Mau B., Glasner J.D.,
RA Rose D.J., Mayhew G.F., Evans P.S., Gregor J., Kirkpatrick H.A.,
RA Postel G., Hackett J., Klink S., Boutin A., Shao Y., Miller L.,
RA Grotbeck E.J., Davis N.W., Lim A., Dimalanta E.T., Potamoukis K.,
RA Apodaca J., Anantharaman T.S., Lin J., Yen G., Schwartz D.C.,
RA Welch R.A., Blattner F.R.;
RT "Genome sequence of enterohaemorrhagic Escherichia coli O157:H7.";
RL Nature 409:529-533 (2001).

```

```

RN SEQUENCE FROM N.A.
RP STRAIN=O157:H7 / RMD 050952;
RX MEDLINE=21156231; PubMed=11258796;
RA Hayashi T., Makino K., Onishi M., Kurokawa K., Ishii K., Yokoyama K.,
RA Han C.-G., Ohtsubo E., Nakayama K., Murata T., Tanaka M., Tobe T.,
RA Iida T., Takami H., Honda T., Sasakawa C., Ogasawara N., Yasunaga T.,
RA Kuhara S., Shiba T., Hattori M., Shinagawa H.;
RT "Complete genome sequence of enterohaemorrhagic Escherichia coli
RT O157:H7 and genomic comparison with a laboratory strain K-12.";
RL DNA Res. 8:11-22 (2001).
CC -!- FUNCTION: CURLIN IS THE STRUCTURAL SUBUNIT OF THE CURLI. CURLI ARE
CC COILED SURFACE STRUCTURES THAT ASSEMBLE PREFERENTIALLY AT GROWTH
CC TEMPERATURES BELOW 37 DEGREES CELSIUS. CURLI CAN BIND TO
CC FIBRONECTIN.
CC -!- SIMILARITY: BELONGS TO THE CSGA/CSGB FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; AF275733; AA53212.1; -
CC EMBL; AE005315; AAG55788.1; -
CC EMBL; AP002554; BAE34843.1; -
CC PIR; D90806; D90806.
CC PIR; H85665; H85665.
CC Fimbria; Signal; Complete proteome.
CC SIGNAL 1 20 BY SIMILARITY.
CC CHAIN 21 152 MAJOR CURLIN SUBUNIT.
CC SEQUENCE 152 AA; 15099 MW; BE2D2D94DDE91243 CRC64;

Query Match 64.4%; Score 499.5; DB 1; Length 152;
Best Local Similarity 67.8%; Pred. No. 3.8e-35;
Matches 103; Conservative 18; Mismatches 30; Indels 1; Gaps 1;

Qy 1 MKLLKVAAPAAIVVSGSALAGVVPQWGGGNGHNGGNSGPDSTLSIYQYGSAALAL 59
Db 1 MKLLKVAALAAIVFSGSALAGVVPQYGGGNGHNGGNSGPNSELNIYQYGGNSALAL 60
Qy 60 QSDARKSETTITQSGYGNAGDVGGADNYDQLVTRVVTTHMAHADOWNAKNSDITVQYGG 119
Db 61 QADARNSDLITQHGCGNGADVGGSDSSIDLQTFGFGNSATLDQWNGKNSMTVKQFG 120
Qy 120 GNGAALVNOTASDSVMVQVGFNNATANOY 151
Db 121 GNGAAVDQTASNSVNVTVQVGFNNATAHOY 152

RESULT 4
ICEN_XANCT STANDARD; PRT; 1567 AA.
AC P18127;
DT 01-NOV-1990 (Rel. 16, Created)
DT 01-NOV-1990 (Rel. 16, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Ice nucleation protein.
GN INAX.
OS Xanthomonas campestris (pv. translucens).
OC Bacteria; Proteobacteria; Gammaproteobacteria; Xanthomonadales;
OC Xanthomonadaceae; Xanthomonas.
OX NCBI_TaxID=343;
RN SEQUENCE FROM N.A.
RP STRAIN=X356S;
RX MEDLINE=91080859; PubMed=2259339;
RA Zhao J., Orser C.S.;
RT "Conserved repetition in the ice nucleation gene inax from
RT Xanthomonas campestris pv. translucens.";
RL Mol. Gen. Genet. 223:163-166 (1990).

```



```

DR EcoGene; EG12621; cs9B.
KW Fimbria; Signal; Complete proteome.
FT SIGNAL 1 21 POTENTIAL.
FT CHAIN 22 151 MINOR CURLIN SUBUNIT.
SQ SEQUENCE 151 AA; 15882 MW; B18D266B964014B8 CRC64;

Query Match
Best Local Similarity 28.1%; Pred. No. 0.36; Indels 5; Gaps 2;
Matches 32; Conservative 15; Mismatches 62;

QY 38 SSGPSTLSIYQGSANALALQSPARKSETTITQSGYGNAGDVQGGADNYDQLVTRVVT 97
Db 21 AAGYDLANSEYNF-----AVNELSKSFNQAALIGAGTNSAQLRQGGSKLLAVVAQSGS 76

QY 98 HEMAHADQNAKNSDITVQYCGNNAALVNQTPASDSSVWVQVGFNNATANOY 151
Db 77 SNRAKIDGTGYNL-AYIDQASANDASTSQAGYNTAMIIQKSGNKANITQY 129

RESULT 6
ICEN_PANAN STANDARD; PRT; 1034 AA.
AC Q47879;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Ice nucleation protein Inau.
GN INAU.
OS Pantoea ananas (Erwinia uredovora).
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Pantoea.
OX NCBI_TaxID=553;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=KUIN-3;
RC MEDLINE=94264407; PubMed=7764866;
RA Michigami Y., Watabe S., Abe K., Obata H., Arai S.;
RT "Cloning and sequencing of an ice nucleation active gene of Erwinia
RT uredovora";
RL Biosci. Biotechnol. Biochem. 58:762-764(1994).
CC -!- FUNCTION: Ice nucleation proteins enable bacteria to nucleate
CC crystallization in supercooled water.
CC -!- SUBCELLULAR LOCATION: Outer membrane.
CC -!- DOMAIN: CONTAINS IMPERFECT REPEATS OF A CONSENSUS OCTAPEPTIDE
CC A-G-Y-G-S-T-X-T; FURTHER ON A 16-RESIDUE AND A REGIONAL 48-RESIDUE
CC PERIODICITY IS SUPERIMPOSED.
CC -!- MISCELLANEOUS: A STRUCTURAL MODEL IS SUGGESTED IN WHICH THE ICE
CC NUCLEATION PROTEIN DISPLAYS A SYMMETRY RELATED TO THAT OF ICE.
CC -!- SIMILARITY: Belongs to the bacterial ice nucleation protein
CC family.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; D14992; BAA03636.1; --
CC PIR; JC2143; JC2143.
CC HSP; P06620; LINA.
CC Pfam; PF00818; Ice_nucleatn.
CC PRINTS; PR00327; ICENUCLEATN.
CC PROSITE; PS00314; ICE_NUCLEATION; 34.
KW Ice nucleation; Repeat; Outer membrane.
FT DOMAIN 162 993 OCTAPEPTIDE PERIODICITY.
SQ SEQUENCE 1034 AA; 103378 MW; FA222523D333ADD CRC64;

Query Match
Best Local Similarity 11.8%; Score 91.5; DB 1; Length 1034;
Matches 46; Conservative 20; Mismatches 54; Indels 55; Gaps 10;

QY 12 IVVSGSALAGVVPQW--GGGNGHNGGNS--GPDSTLSIYQGSANAL 57
Db 161 IATYGSTLSGTHQSQLIAGTGSTETAGSDSTLIAGYGTGTAGSDSTL-VAGYGSTQTA- 218

QY 58 ALOSDARKSETTITQSGYGN-----GADVQGG-----ADNYDQLVTRVVTHEMAHAD- 104
Db 219 -----GEESQAGYGTGTGMKGSDLTACGTAGTGTAGDDSSLIAGYGTGTAGSDS 270

QY 105 -----OWNAKNSDITVQYCGNNAALVNQTPASDSSVWVQVGFNNATANOQ 150
Db 271 SLTAGYGSTQTAQKGSDLTAG-YGSTGTA-----GADSSLI-----AGYGSTQTAGE 316

RESULT 7
ICEN_ERWHE STANDARD; PRT; 1258 AA.
AC P16239;
DT 01-APR-1990 (Rel. 14, Created)
DT 01-APR-1990 (Rel. 14, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Ice nucleation protein.
GN ICE.
OS Erwinia herbicola.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Pantoea.
OX NCBI_TaxID=549;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=M1;
RC MEDLINE=90152370; PubMed=25151997;
RA Warren G.J., Corotto L.V.;
RT "The consensus sequence of ice nucleation proteins from Erwinia
RT herbicola, Pseudomonas fluorescens and Pseudomonas syringae.";
RL Gene 85:239-242(1989).
CC -!- FUNCTION: Ice nucleation proteins enable bacteria to nucleate
CC crystallization in supercooled water.
CC -!- SUBCELLULAR LOCATION: Outer membrane.
CC -!- DOMAIN: CONTAINS 126 IMPERFECT REPEATS OF A CONSENSUS OCTAPEPTIDE
CC A-G-Y-G-S-T-X-T; FURTHER ON A 16-RESIDUE AND A REGIONAL 48-RESIDUE
CC PERIODICITY IS SUPERIMPOSED.
CC -!- MISCELLANEOUS: A STRUCTURAL MODEL IS SUGGESTED IN WHICH THE ICE
CC NUCLEATION PROTEIN DISPLAYS A SYMMETRY RELATED TO THAT OF ICE.
CC -!- SIMILARITY: Belongs to the bacterial ice nucleation protein
CC family.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; M26382; AAA24823.1; --
CC PIR; JQ0188; JQ0188.
CC HSP; P06620; LINA.
CC InterPro; IPR000258; Ice_nucleatn.
CC Pfam; PF00818; Ice_nucleation; 65.
CC PRINTS; PR00327; ICENUCLEATN.
CC PROSITE; PS00314; ICE_NUCLEATION; 45.
KW Ice nucleation; Repeat; Outer membrane.
FT DOMAIN 162 1217 OCTAPEPTIDE PERIODICITY.
SQ SEQUENCE 1258 AA; 125084 MW; 590E8A130077FBD4 CRC64;

Query Match
Best Local Similarity 11.8%; Score 91.5; DB 1; Length 1258;
Matches 46; Conservative 20; Mismatches 54; Indels 55; Gaps 10;

QY 12 IVVSGSALAGVVPQW--GGGNGHNGGNS--GPDSTLSIYQGSANAL 57
Db 161 IATYGSTLSGTHQSQLIAGTGSTETAGSDSTLIAGYGTGTAGADSTL-VAGYGSTQTA- 218

```

```

QY 58 ALQSDARKSETTITOGYGN-----GADVGG-----ADNYDQIVTRVWTHMAHAD- 104
Db 219 -----GEESQAGYSTGTGKMSDLTAGYSTGTAGDDSSLIAGYGSTQTAGEDS 270
QY 105 -----QWNAKNSDITVQYGGNNAALVNOTASDSSVMVRQVGFGNNAANQ 150
Db 271 SLTAGYGSTQTAGKMSDLTAG-YGSTGTA-----GADSSLI---AGYGSTOTAGE 316

RESULT 8
PMPB_CHLTR STANDARD; PRT; 1754 AA.
ID PMPB_CHLTR
AC 084418;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Probable outer membrane protein pmpB precursor (Polymorphic membrane protein B).
DE PMPB OR CT413.
OS Chlamydia trachomatis.
OC Bacteria; Chlamydiae; Chlamydiales; Chlamydiaceae; Chlamydia.
OX NCBI_TaxID=813;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=D/UW-3/Cx;
RX MEDLINE=99000809; PubMed=9784136;
RA Stephens R.S., Kallan S., Lammel C.J., Fan J., Marathe R., Aravind L., Mitchell W.P., Olinger L., Tatusov R.L., Zhao Q., Koonin E.V., Davis R.W.;
RA "Genome sequence of an obligate intracellular pathogen of humans: Chlamydia trachomatis."
RT Science 282:754-759 (1998).
RL
CC -1- SUBCELLULAR LOCATION: Cell wall surface (elementary bodies) (Potential).
CC -1- SIMILARITY: Belongs to the PMP outer membrane protein family.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See http://www.isb-sib.ch/announce/ or send an email to license@isb-sib.ch).
CC
CC EMBL; AB001314; AAC68010.1; ALT_INIT.
CC PHCI-2DPAGE; 084418; -.
CC InterPro; IPR003368; Chlamydia_PMP.
CC Pfam; PF02415; Chlamydia_PMP; 4.
CC TIGRPFAMs; TIGR01376; POMP repeat; 18.
KW Outer membrane; Signal; Multigene family; Complete proteome.
FT SIGNAL 1 14 POTENTIAL.
FT CHAIN 15 1754 PROBABLE OUTER MEMBRANE PROTEIN PMPB.
SQ SEQUENCE 1754 AA; 183317 MW; 969CF8D85D36185D CRC64;

Query Match 11.7%; Score 91; DB 1; Length 1754;
Best Local Similarity 33.1%; Pred. No. 7.3;
Matches 40; Conservative 10; Mismatches 59; Indels 12; Gaps 5;

QY 15 SCSALAGVVPWCGGNG--HNGGNSGDPSTLSIYQYGSNAALALQSDARKSETTIT- 71
Db 397 NADAWASSPQSGSGATTVNSGDSGSDSTSETVPATKGG-GLYTDKNLSITNIG 455
QY 72 --QSGYGNGADVQGGADNDQIVTRVWTHMAHADQWNAKNSDITVGO-YGGNNAALVNQ 128
Db 456 IIEIANKATDVGGGA-----YKGTLTCTENSHRLQFLKNSDKQGGIYGEDNITLNL 510
QY 129 T 129
Db 511 T 511

RESULT 9
CSGB_SALTY STANDARD; PRT; 151 AA.
ID CSGB_SALTY
AC 0827M3;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Minor curlin subunit precursor.
DE CSGB OR STY1180 OR T1777.
OS Salmonella typhi.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales; Enterobacteriaceae; Salmonella.
OX NCBI_TaxID=601;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CT18;
RX MEDLINE=21534947; PubMed=11677608;
RA Parkhill J., Dougan G., James K.D., Thomson N.R., Pickard D., Wain J., Churcher C., Mungall K.L., Bentley S.D., Holden M.T.G., Sebahia M., Baker S., Basham D., Brooks K., Chillingworth T., Connor P., Cronin A., Davis P., Davies R.M., Dowd L., White N., Farrar J., Feitwell T., Hamlin N., Haque A., Hien T.T., Holroyd S., Jagels K., Krogh A., Larsen T.S., Leather S., Moule S., O'Gaora P., Parry C., Quail M.A., Rutherford K., Simmonds M., Skelton J., Stevens K., Whitehead S., Barrall B.G.;
RA "Complete genome sequence of a multiple drug resistant Salmonella enterica serovar Typhi CT18."
RT Nature 413:848-852 (2001).
RL
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=Ty2 / ATCC 700931;
RX MEDLINE=22531367; PubMed=12644504;
RA Deng W., Liou S.-R., Plunkett G. III, Mayhew G.F., Rose D.J., Burland V., Kodyianni V., Schwartz D.C., Blattner F.R.;
RA "Comparative genomics of Salmonella enterica serovar Typhi strains Ty2 and CT18."
RL J. Bacteriol. 185:2330-2337 (2003).
CC -1- FUNCTION: CURLIN IS THE STRUCTURAL SUBUNIT OF THE CURLI. CURLI ARE COILED SURFACE STRUCTURES THAT ASSEMBLE PREFERENTIALLY AT GROWTH TEMPERATURES BELOW 37 DEGREES CELSIUS. CURLI CAN BIND TO FIBRONECTIN. THE MINOR SUBUNIT IS THE NUCLEATION COMPONENT OF CURLIN MONOMERS.
CC -1- SIMILARITY: BELONGS TO THE CSGB/CSGB FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See http://www.isb-sib.ch/announce/ or send an email to license@isb-sib.ch).
CC
CC EMBL; AL627269; CAD08267.1; -.
CC EMBL; AE016840; AAO69400.1; -.
KW Fimbria; Signal; Complete proteome.
FT SIGNAL 1 21 POTENTIAL.
FT CHAIN 22 151 MINOR CURLIN SUBUNIT.
SQ SEQUENCE 151 AA; 16254 MW; 161C54326E573495 CRC64;

Query Match 11.7%; Score 90.5; DB 1; Length 151;
Best Local Similarity 28.8%; Pred. No. 0.53;
Matches 34; Conservative 19; Mismatches 46; Indels 19; Gaps 6;

QY 51 GSANAALALQSDARKSE-----TTITQSGYNGADVGO-GADNYDQI-----VTRVWTH 98
Db 14 GAPCIATATNYDLARGEYNFAVNELSKSPNQAAIIQGVGTDSARVROBGSKLLSVISQ 73
QY 99 E-----MAHADQWNAKNSDIT-VGOYGGNNAALVNQTSDDSSVMVRQVGFGNNAANQ 151
Db 74 EGENNRKAVDQ--AGNYNFAYIEQTGNANDASISQAYGNSAAIIQKSGNKNITQY 129

RESULT 10
CSGB_SALTY STANDARD; PRT; 151 AA.
ID CSGB_SALTY

```

AC P52226;  
DT 01-OCT-1996 (Rel. 34, Created)  
DT 01-OCT-1996 (Rel. 34, Last sequence update)  
DT 28-FEB-2003 (Rel. 41, Last annotation update)  
DE Minor curlin subunit precursor (Pimbrin SEF17 minor subunit).  
CSGB OR AGFB OR STM1143, and  
GN Salmonella typhimurium, and  
OS Salmonella enteritidis.  
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;  
OC Enterobacteriaceae; Salmonella.  
ON NCBI TaxID=602, 592;  
OX (1)  
RP SEQUENCE FROM N.A.  
RC SPECIES=S.typhimurium; STRAIN=SR-11;  
RX MEDLINE=98117058; PubMed=9457880;  
RA Romling U., Bian Z., Hammar M., Sierralta W.D., Normark S.;  
RT "Curli fibers are highly conserved between Salmonella typhimurium and  
RT Escherichia coli with respect to operon structure and regulation.";  
RL J. Bacteriol. 180:722-731(1998).  
RN (2)  
RP SEQUENCE FROM N.A.  
RC SPECIES=S.typhimurium; STRAIN=LT2 / SGSC1412 / ATCC 700720;  
RX MEDLINE=21534948; PubMed=11677609;  
RA McClelland M., Sanderson K.E., Spieth J., Clifton S.W., Latreille P.,  
RA Courtney L., Porwollik S., Ali J., Dante M., Du F., Hou S., Layman D.,  
RA Leonard S., Nguyen C., Scott K., Holmes A., Grewal N., Mulvaney E.,  
RA Ryan E., Sun H., Florea L., Miller W., Stoneking T., Nhan M.,  
RA Waterston R., Wilson R.K.;  
RT "Complete genome sequence of Salmonella enterica serovar Typhimurium  
RT LT2.";  
RL Nature 413:852-856(2001).  
RN (3)  
RP SEQUENCE FROM N.A.  
RC SPECIES=S.enteritidis; STRAIN=27655-3B;  
RX MEDLINE=96146512; PubMed=8550497;  
RA Collinson S.K., Clouthier S.C., Doran J.L., Banser P.A., Kay W.W.;  
RT "Salmonella enteritidis ageBAC operon encoding thin, aggregative  
RT fimbriae";  
RL J. Bacteriol. 178:662-667(1996).  
CC -!- FUNCTION: CURLIN IS THE STRUCTURAL SUBUNIT OF THE CURLI ARE  
CC COILED SURFACE STRUCTURES THAT ASSEMBLE PREFERENTIALLY AT GROWTH  
CC TEMPERATURES BELOW 37 DEGREES CELSIUS. CURLI CAN BIND TO  
CC FIBRONECTIN. THE MINOR SUBUNIT IS THE NUCLEATION COMPONENT OF  
CC CURLIN MONOMERS.  
CC -!- SIMILARITY: BELONGS TO THE CSGA/CSGB FAMILY.  
CC  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC  
CC EMBL; AJ002301; CAA05316.1; -  
CC EMBL; AE008749; AAL20073.1; -  
CC EMBL; U43280; AAC43598.1; -  
CC PIR; JC6040; JC6040.  
CC StyGene; SGI0609; CSGB.  
KW Fimbria; Signal; Complete proteome.  
FT SIGNAL 1 21 POTENTIAL.  
FT CHAIN 22 151 MINOR CURLIN SUBUNIT.  
SQ SEQUENCE 151 AA; 16182 MW; 16182 MW; CQFC5430E6DD361D CRC64;  
Query Match 11.7%; Score 90.5; DB 1; Length 151;  
Best Local Similarity 27.5%; Pred. No. 0.53;  
Matches 36; Conservative 17; Mismatches 39; Indels 39; Gaps 5;  
OY 10 AAIVSGSAGAGVPPQWGGGNGHNGGNSGPDSTLSIYQYGSNAAALQSDARKSETT 69  
DB 58 ARVQEGSKLSVISQ--EGGNRAKVDQAGNYFAYIEQTGNAN-----DAS 103  
OY 70 ITQSGYGNAGDVGO--GADNYQLVTRVVTHEMAHADQWNAKNSDITVQYGGNNAALVQ 128

DB 104 ISQAYGNSAAIIQKSGN-----KANIT--QYGTOKTAAVVQ 139  
OY 129 TASDSSVMVQ 139  
DB 140 QKSEHAIKRVQ 150  
RESULT 11  
ICEA PANAN  
ID ICEA PANAN STANDARD; PRT; 1322 AA.  
AC P20459;  
DT 01-FEB-1991 (Rel. 17, Created)  
DT 01-FEB-1991 (Rel. 17, Last sequence update)  
DT 16-OCT-2001 (Rel. 40, Last annotation update)  
DE Ice nucleation protein inaa.  
GN INAA.  
OS Pantoea ananas (Erwinia uredovora).  
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;  
OC Enterobacteriaceae; Pantoea.  
OX NCBI TaxID=553;  
RN (1)  
RP SEQUENCE FROM N.A.  
RX MEDLINE=90092494; PubMed=2599095;  
RA Abe K., Watabe S., Emori Y., Watanabe M., Arai S.;  
RT "An ice nucleation active gene of Erwinia ananas. Sequence similarity  
RT to those of Pseudomonas species and regions required for ice  
RT nucleation activity.";  
RL FEBS Lett. 258:297-300(1989).  
CC -!- FUNCTION: Ice nucleation proteins enable bacteria to nucleate  
CC crystallization in supercooled water.  
CC -!- SUBCELLULAR LOCATION: Outer membrane (By similarity).  
CC -!- DOMAIN: CONTAINS MANY IMPERFECT REPEATS OF THE CONSENSUS  
CC OCTAPEPTIDE A-G-Y-G-S-T-X-T; FURTHER ON A 16-RESIDUE AND A  
CC REGIONAL 48-RESIDUE PERIODICITY IS SUPERIMPOSED.  
CC -!- SIMILARITY: Belongs to the bacterial ice nucleation protein  
CC family.  
CC  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC  
CC EMBL; X17316; CAA35194.1; -  
CC PIR; S07053; S07053.  
CC HSSP; P06620; 11NA.  
CC InterPro; IPR000258; Ice nucleatn.  
CC Pfam; PF00818; Ice nucleation; 69.  
CC PRINTS; PR00327; ICENUCLEATN.  
CC PROSITE; PS00314; ICE\_NUCLEATION; 49.  
KW Ice nucleation; Repeat; Outer membrane.  
FT DOMAIN 162 1281 OCTAPEPTIDE PERIODICITY.  
SQ SEQUENCE 1322 AA; 131094 MW; 89B0BE24AA837039 CRC64;  
Query Match 11.5%; Score 89.5; DB 1; Length 1322;  
Best Local Similarity 26.8%; Pred. No. 7.1;  
Matches 34; Conservative 18; Mismatches 26; Indels 49; Gaps 7;  
OY 34 GGNSSGPDSTLSIYQYGSNAAALQSDARKSETTITQSGYGNAGDVGGADNYQLVT 93  
DB 933 GSTTAGPDSSL-IAGYSTQTA-----GYSILTAYG- 965  
OY 94 RVVTHEMAHADQWNAKNSDITVQYG-----GNNAALV-----NOTASDSSVMVQVGF 143  
DB 966 -----STQTSQNSDLITG-YGSTSTAGYESSLIAGYSTQTASFKSLM--AGY 1013  
OY 144 NNATANQ 150  
DB 1014 SSQTARE 1020

RESULT 12  
 YK98 MYCTU  
 ID\_YK98\_MYCTU STANDARD; PRT; 491 AA.  
 AC Q10707;  
 DT 01-OCT-1996 (Rel. 34, Created)  
 DT 16-OCT-2001 (Rel. 40, Last sequence update)  
 DT 10-OCT-2003 (Rel. 42, Last annotation update)  
 DE Hypothetical glycine-rich protein RV2098c/MT2159/Mb2125c.  
 GN RV2098C OR MT2159 OR MTCV49.38C OR MB2125C.  
 OS Mycobacterium tuberculosis, and  
 OS Mycobacterium bovis.  
 OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;  
 OC Corynebacterineae; Mycobacteriaceae; Mycobacterium.  
 OX NCBI\_TaxID=1773, 1765;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC SPECIES=M.tuberculosis; STRAIN=H37RV;  
 RX MEDLINE=98295987; PubMed=9634230;  
 RA Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,  
 RA Gordon S.V., Eiglmeier K., Gas S., Barry C.E. III, Tekala F.,  
 RA Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,  
 RA Davies R., Devlin K., Feltwell T., Gentles S., Hamlin N., Holroyd S.,  
 RA Hornsby T., Jagels K., Krogh A., McLean J., Moule S., Murphy L.,  
 RA Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,  
 RA Rutter S., Seeger K., Skelton S., Squares S., Squares R.,  
 RA Sulston J.E., Taylor K., Whitehead S., Barrall B.G.;  
 RA "Deciphering the biology of Mycobacterium tuberculosis from the  
 RT complete genome sequence."  
 RL Nature 393:537-544(1998).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC SPECIES=M.tuberculosis; STRAIN=CDC 1551 / Oshkosh;  
 RX MEDLINE=22206494; PubMed=12218036;  
 RA Fleischmann R.D., Alland D., Eisen J.A., Carpenter L., White C.,  
 RA Peterson J., DeBoy R., Dodson R., Gwinn M., Haft D., Hickey E.,  
 RA Kolonay J.F., Nelson W.C., Umayam L.A., Ermolaeva M., Salzberg S.L.,  
 RA Delcher A., Utterback T., Weidman J., Khouri H., Gill J., Mikula A.,  
 RA Bishai W., Jacobs W.R. Jr., Venter J.C., Fraser C.M.;  
 RA "Whole-genome comparison of Mycobacterium tuberculosis clinical and  
 RT laboratory strains."  
 RL J. Bacteriol. 184:5479-5490(2002).  
 RN [3]  
 RP SEQUENCE FROM N.A.  
 RC SPECIES=M.bovis; STRAIN=AF2122/97;  
 RX MEDLINE=22709107; PubMed=12788972;  
 RA Garnier T., Eiglmeier K., Camus J.-C., Medina N., Mansoor H.,  
 RA Pryor M., Duthoy S., Grondin S., Lacroix C., Monsemp C., Simon S.,  
 RA Harris B., Atkin R., Doggett J., Mayes R., Keating L., Wheeler P.R.,  
 RA Parkhill J., Barrall B.G., Cole S.T., Gordon S.V., Hewinson R.G.;  
 RA "The complete genome sequence of Mycobacterium bovis."  
 RL Proc. Natl. Acad. Sci. U.S.A. 100:7877-7882(2003).  
 CC -!- SIMILARITY: BELONGS TO THE MYCOBACTERIAL PE FAMILY. PGRS  
 CC -!- SUBFAMILY:  
 CC -!- CAUTION: Ref.1 sequence differs from that shown due to a  
 CC frameshift in position 59. Ref.1 sequence has been checked by  
 CC authors in Ref.1 and they report that no errors have been found.  
 CC  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL Outstation  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC  
 CC EMBL; Z73966; CAB98228.1; ALT\_FRAME.  
 CC EMBL; AB007065; AKA46440.1; -;  
 CC EMBL; XZ48341; CAD96978.1; -;  
 CC TIGR; MT2159; -;  
 CC TubercuList; RV2098c; -;  
 CC InterPro; IPR000084; PE\_region.  
 CC Pfam; PF00934; PE; 1.  
 CC  
 CC ProDom; PD001223; PE region; 1.  
 KW Hypothetical protein; Complete proteome.  
 FT CONFLICT 312 G -> GG (IN REF. 1).  
 SQ SEQUENCE 491 AA; 41979 MW; 12C8630C59CA0C13 CRC64;  
 Query Match 11.2%; Score 87; DB 1; Length 491;  
 Best Local Similarity 28.2%; Pred. No. 3.8; 45; Indels 26; Gaps 4;  
 Matches 31; Conservative 8; Mismatches 45; Indels 26; Gaps 4;  
 QY 20 AGIVPQWCGGNN-----HNGGNSGPDSTLSITVQYSANALALQSDARKSETTITQSG 74  
 DB 222 AGLIHGGAGNGGDDGGHGGSGKAGSGGGGFGQFGAGGLL----- 264  
 QY 75 YNGADVGGGADNDYDLVTRVTHEMAHADOWNAKNSD---ITVGQYGGN 121  
 DB 265 YNGGAGSGGNGGD-AGTGVSSDGFAGLGGSGGSGGAGLIGVGGGGN 313  
 RESULT 13  
 GP63 LEIDO  
 ID GP63 LEIDO STANDARD; PRT; 590 AA.  
 AC P23223;  
 DT 01-NOV-1991 (Rel. 20, Created)  
 DT 01-NOV-1991 (Rel. 20, Last sequence update)  
 DT 10-OCT-2003 (Rel. 42, Last annotation update)  
 DE Leishmanolysin precursor (EC 3.4.24.36) (Cell surface protease)  
 DE (Major surface glycoprotein) (GP63 protein) (Promastigote surface  
 DE endopeptidase).  
 GN GP63.  
 OS Leishmania donovani.  
 OS Eukaryota; Euglenozoa; Kinetoplastida; Trypanosomatidae; Leishmania.  
 OX NCBI\_TaxID=5661;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=LV9;  
 RX MEDLINE=92107220; PubMed=1762629;  
 RA Webb J.R., Button L.L., McMaster R.W.;  
 RA "Heterogeneity of the genes encoding the major surface glycoprotein  
 RT of Leishmania donovani."  
 RL Mol. Biochem. Parasitol. 48:173-184(1991).  
 CC -!- FUNCTION: Has an integral role during the infection of macrophages  
 CC in the mammalian host.  
 CC -!- CATALYTIC ACTIVITY: Preference for hydrophobic residues at P1 and  
 CC P1' and basic residues at P2 and P3'. A model nonapeptide is  
 CC cleaved at Ala-Tyr-|-Leu-Lys-Lys-  
 CC -!- COFACTOR: Binds 1 zinc ion per subunit (By similarity).  
 CC -!- SUBCELLULAR LOCATION: Attached to the membrane by a GPI-anchor.  
 CC -!- SIMILARITY: Belongs to peptidase family M8.  
 CC  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL Outstation  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC  
 CC EMBL; M60048; AAK29244.1; -;  
 CC HESP; P08148; LML.  
 CC MEROPS; M08.001; -;  
 CC InterPro; IPR006025; Pept\_Mn\_Zn\_BS.  
 CC InterPro; IPR001577; Peptidase\_M8.  
 CC Pfam; PF01457; Peptidase\_M8; 1-  
 CC PRINTS; PR00782; ISHMANOLYSIN.  
 CC PROSITE; PS00142; ZINC\_PROTEASE; 1  
 CC Hydrolase; Metalloprotease; Glycoprotein; Metal-binding; Zinc;  
 CC Zymogen; Signal; Cell adhesion; GPI-anchor; Lipoprotein.  
 CC SIGNAL 1 39 POTENTIAL.  
 CC PROPEP 40 87 ACTIVATION PEPTIDE.  
 CC CHAIN 88 565 LEISHMANOLYSIN.  
 CC PROPEP 566 590 REMOVED IN NATURE FORM (BY SIMILARITY).  
 CC METAL 251 251 ZINC (CATALYTIC) (BY SIMILARITY).  
 CC ACT\_SITE 252 252 BY SIMILARITY.



```

FT METAL 255 255 ZINC (CATALYTIC) (BY SIMILARITY).
FT METAL 321 321 ZINC (CATALYTIC) (BY SIMILARITY).
FT DISULFID 112 129 BY SIMILARITY.
FT DISULFID 178 217 BY SIMILARITY.
FT DISULFID 301 373 BY SIMILARITY.
FT DISULFID 380 443 BY SIMILARITY.
FT DISULFID 393 412 BY SIMILARITY.
FT DISULFID 402 477 BY SIMILARITY.
FT DISULFID 454 498 BY SIMILARITY.
FT DISULFID 503 553 BY SIMILARITY.
FT DISULFID 523 546 BY SIMILARITY.
FT CARBOHYD 287 287 N-LINKED (GLCNAC. .) (POTENTIAL).
FT LIPID 565 565 GPI-anchor amidated asparagine (By similarity).
SQ SEQUENCE 590 AA; 62950 MW; 0FB315D299659F58 CRC64;

Query Match 11.2%; Score 87; DB 1; Length 590;
Best Local Similarity 89.5%; Pred. No. 4.7;
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 85 ADNYDQLVTRVVTHEMAHA 103
DB 238 ASRYDQLVTRVVTHEMAHA 256

RESULT 14
GP63 LEICH
ID GP63 LEICH STANDARD; PRT; 599 AA.
AC P15706;
DT 01-APR-1990 (Rel. 14, Created)
DT 01-APR-1990 (Rel. 14, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Leishmanolysin precursor (EC 3.4.24.36) (Cell surface protease)
DE (Major surface glycoprotein) (GP63 protein) (Promastigote surface endopeptidase).
GN GP63.
OS Leishmania chagasi.
OC Eukaryota; Euglenozoa; Kinetoplastida; Trypanosomatidae; Leishmania.
OX NCBI_TaxID=44271;
RN [1]
RP SEQUENCE FROM N.A. PubMed=3320059;
RX MEDLINE=90205976; PubMed=3320059;
RA Ramamoorthy R., Donelson J.E., Paetz K.E., Maybodi M., Roberts S.C., Wilson M.E.;
RT "Leishmania gp63 molecule implicated in cellular adhesion lacks an Arg-Gly-Asp sequence.";
RL Mol. Biochem. Parasitol. 39:267-274 (1990).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=92112918; PubMed=1370484;
RA Ramamoorthy R., Donelson J.E., Paetz K.E., Maybodi M., Roberts S.C., Wilson M.E.;
RT "Three distinct RNAs for the surface protease gp63 are differentially expressed during development of Leishmania donovani chagasi promastigotes to an infectious form.";
RL J. Biol. Chem. 267:1888-1895 (1992).
CC -!- FUNCTION: Has an integral role during the infection of macrophages in the mammalian host.
CC -!- CATALYTIC ACTIVITY: Preference for hydrophobic residues at P1 and P1' and basic residues at P2 and P3'. A model nonapeptide is cleaved at -Ala-Tyr-|-Leu-Lys-Lys-.
CC -!- COFACTOR: Binds 1 zinc ion per subunit (By similarity).
CC -!- SUBCELLULAR LOCATION: Attached to the membrane by a GPI-anchor.
CC -!- SIMILARITY: Belongs to peptidase family M8.
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See http://www.isb-sib.ch/announce/ or send an email to license@isb-sib.ch).
CC EMBL; M80672; AAA29238.1; -

```

```

DR EMBL; M28527; AAA29235.1; -.
DR PIR; A44951; A44951.
DR HSSP; P08148; 1LML.
DR MEROPS; M08.001; -.
DR InterPro; IPR006025; Pept M Zn BS.
DR Pfam; PF01457; Peptidase_M8; 1.
DR PRINTS; PR00762; LSHMANOLYSIN.
DR PROSITE; PS00142; ZINC_PROTEASE; 1.
KW Hydrolase; Metalloprotease; Glycoprotein; Metal-binding; Zinc;
KW Zymogen; Signal; Cell adhesion; GPI-anchor; Lipoprotein.
FT SIGNAL 1 39 POTENTIAL.
FT PROPEP 40 97 ACTIVATION PEPTIDE.
FT CHAIN 98 574 LEISHMANOLYSIN.
FT PROPEP 575 599 REMOVED IN MATURE FORM (BY SIMILARITY).
FT METAL 261 261 ZINC (CATALYTIC) (BY SIMILARITY).
FT ACT_SITE 262 262 BY SIMILARITY.
FT METAL 331 331 ZINC (CATALYTIC) (BY SIMILARITY).
FT DISULFID 122 139 ZINC (CATALYTIC) (BY SIMILARITY).
FT DISULFID 188 227 BY SIMILARITY.
FT DISULFID 311 383 BY SIMILARITY.
FT DISULFID 390 452 BY SIMILARITY.
FT DISULFID 403 422 BY SIMILARITY.
FT DISULFID 412 486 BY SIMILARITY.
FT DISULFID 463 507 BY SIMILARITY.
FT DISULFID 512 562 BY SIMILARITY.
FT DISULFID 532 555 BY SIMILARITY.
FT CARBOHYD 297 297 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 394 394 N-LINKED (GLCNAC. .) (POTENTIAL).
FT LIPID 574 574 GPI-anchor amidated asparagine (By similarity).
SQ SEQUENCE 599 AA; 63848 MW; 746730AB8E2A2E7C CRC64;

Query Match 11.2%; Score 87; DB 1; Length 599;
Best Local Similarity 89.5%; Pred. No. 4.8;
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 85 ADNYDQLVTRVVTHEMAHA 103
DB 248 ASRYDQLVTRVVTHEMAHA 266

RESULT 15
GP63 LEIMA
ID GP63 LEIMA STANDARD; PRT; 602 AA.
AC P08148; P15906;
DT 01-AUG-1988 (Rel. 08, Created)
DT 01-APR-1990 (Rel. 14, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Leishmanolysin precursor (EC 3.4.24.36) (Cell surface protease)
DE (Major surface glycoprotein) (GP63 protein) (Promastigote surface endopeptidase).
GN GP63.
OS Leishmania major.
OC Eukaryota; Euglenozoa; Kinetoplastida; Trypanosomatidae; Leishmania.
OX NCBI_TaxID=5664;
RN [1]
RP SEQUENCE FROM N.A., AND SEQUENCE OF 101-123.
RX MEDLINE=88154764; PubMed=3346625;
RA Buton L.L., McMaster W.R.;
RT "Molecular cloning of the major surface antigen of leishmania.";
RL J. Exp. Med. 167:724-729 (1988).
RN [2]
RP REVISIONS.
RA Buton L.L., McMaster W.R.;
RL J. Exp. Med. 171:589-589 (1990).
RN [3]
RP GPI-ANCHOR.
RX MEDLINE=91009116; PubMed=2145267;
RA Schneider P., Ferguson M.A.J., McConville M.J., Mehlert A., Homans S.W., Bordier C.;
RT "Structure of the glycosyl-phosphatidylinositol membrane anchor of

```

the Leishmania major promastigote surface protease.";  
 J. Biol. Chem. 265:16955-16964(1990).  
 [4]  
 RX X-RAY CRYSTALLOGRAPHY (2.6 ANGSTROMS).  
 MEDLINE=95406217; PubMed=7675788;  
 RA Schlagenhauf E., Etges R., Metcalf P.;  
 RT "Crystallization and preliminary X-ray diffraction studies of  
 leishmanolysin, the major surface metalloproteinase from Leishmania  
 major.";  
 RT Proteins 22:58-66(1995).  
 RL [5]  
 RN X-RAY CRYSTALLOGRAPHY (1.86 ANGSTROMS).  
 MEDLINE=98416698; PubMed=9739094;  
 RA Schlagenhauf E., Etges R., Metcalf P.;  
 RT "The crystal structure of the Leishmania major surface proteinase  
 leishmanolysin.";  
 RT Structure 6:1035-1046(1998).  
 CC -!- FUNCTION: Has an integral role during the infection of macrophages  
 in the mammalian host.  
 CC -!- CATALYTIC ACTIVITY: Preference for hydrophobic residues at P1 and  
 P1' and basic residues at P2 and P3'. A model nonapeptide is  
 cleaved at -Ala-Tyr-[Leu-Lys-Lys-...  
 CC -!- COFACTOR: Binds 1 zinc ion per subunit.  
 CC -!- SUBCELLULAR LOCATION: Attached to the membrane by a GPI-anchor.  
 CC -!- PTM: THE PHOSPHATIDYLINOSITOL MOIETY OF THE GPI-ANCHOR CONTAINS A  
 FULLY SATURATED, UNBRANCHED 1-O-ALKYL CHAIN (MAINLY C24:0) AND A  
 MIXTURE OF FULLY SATURATED UNBRANCHED 2-O-ACYL CHAINS (C12:0,  
 C14:0, C16:0, AND C18:0).  
 CC -!- SIMILARITY: Belongs to peptidase family M8.  
 CC  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 the European Bioinformatics Institute. There are no restrictions on its  
 use by non-profit institutions as long as its content is in no way  
 modified and this statement is not removed. Usage by and for commercial  
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC  
 CC EMBL; Y00647; CAA68673.1; -;  
 CC PIR; P0221; P0221.  
 CC PDB; 1LMU; 17-SEP-97.  
 CC MEROPS; M08.001;  
 CC InterPro; IPR006025; Pept\_M\_Zn\_BS.  
 CC InterPro; IPR001577; Peptidase\_M8.  
 CC Pfam; PF01457; Peptidase\_M8; 1.  
 CC PRINTS; PR00782; LSHMANOLYSIN.  
 CC PROSITE; PS00142; ZINC\_PROTEASE; 1.  
 CC Hydrolase; Metalloprotease; Glycoprotein; Metal-binding; Zinc;  
 KW Zymogen; Signal; Cell adhesion; GPI-anchor; 3D-structure; Lipoprotein.  
 FT SIGNAL 1 39  
 FT PROPEP 40 100 ACTIVATION PEPTIDE.  
 FT CHAIN 101 577 LEISHMANOLYSIN.  
 FT PROPEP 578 602 REMOVED IN MATURE FORM.  
 FT METAL 264 264 ZINC (CATALYTIC).  
 FT ACT SITE 265 265  
 FT METAL 268 268 ZINC (CATALYTIC).  
 FT METAL 334 334 ZINC (CATALYTIC).  
 FT DISULFID 125 142  
 FT DISULFID 191 230  
 FT DISULFID 314 386  
 FT DISULFID 393 455  
 FT DISULFID 406 425  
 FT DISULFID 415 489  
 FT DISULFID 466 510  
 FT DISULFID 515 565  
 FT DISULFID 535 558  
 FT CARBOHYD 300 300  
 FT CARBOHYD 407 407  
 FT LIPID 577 577  
 FT STRAND 101 102  
 FT STRAND 107 108  
 FT STRAND 111 114  
 FT HELIX 116 119  
 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 GPI-anchor amidated asparagine.

FT TURN 121 121  
 FT TURN 128 128  
 FT STRAND 131 131  
 FT STRAND 133 133  
 FT STRAND 139 141  
 FT HELIX 144 146  
 FT HELIX 150 158  
 FT TURN 159 159  
 FT HELIX 160 169  
 FT TURN 170 171  
 FT STRAND 172 174  
 FT STRAND 177 178  
 FT STRAND 180 181  
 FT TURN 189 190  
 FT HELIX 191 193  
 FT HELIX 198 202  
 FT TURN 203 203  
 FT STRAND 205 206  
 FT STRAND 210 215  
 FT TURN 221 222  
 FT STRAND 226 232  
 FT TURN 234 235  
 FT STRAND 238 244  
 FT HELIX 247 249  
 FT HELIX 256 269  
 FT TURN 270 271  
 FT HELIX 274 279  
 FT TURN 280 281  
 FT STRAND 283 286  
 FT HELIX 289 291  
 FT STRAND 296 299  
 FT HELIX 302 312  
 FT TURN 313 313  
 FT TURN 315 316  
 FT TURN 320 322  
 FT TURN 328 332  
 FT STRAND 335 335  
 FT TURN 337 339  
 FT TURN 341 342  
 FT STRAND 343 343  
 FT TURN 344 345  
 FT STRAND 353 353  
 FT HELIX 356 364  
 FT TURN 365 366  
 FT STRAND 369 370  
 FT HELIX 372 374  
 FT TURN 380 383  
 FT HELIX 386 390  
 FT STRAND 394 395  
 FT TURN 396 397  
 FT STRAND 398 399  
 FT TURN 402 404  
 FT STRAND 413 414  
 FT TURN 417 418  
 FT STRAND 421 425  
 FT STRAND 428 429  
 FT HELIX 435 437  
 FT TURN 443 444  
 FT STRAND 445 446  
 FT TURN 450 454  
 FT STRAND 458 465  
 FT TURN 466 467  
 FT HELIX 470 472  
 FT TURN 475 477  
 FT HELIX 478 480  
 FT TURN 485 486  
 FT STRAND 487 494  
 FT STRAND 496 496  
 FT STRAND 506 516  
 FT TURN 517 520  
 FT STRAND 521 525  
 FT TURN 527 528  
 FT STRAND 533 534  
 FT TURN 537 538

FT STRAND 540 542  
FT HELIX 543 545  
FT TURN 546 546  
FT STRAND 550 550  
FT TURN 552 553  
FT STRAND 555 557  
FT HELIX 561 565  
FT TURN 566 567  
FT HELIX 569 572  
FT TURN 573 573  
SQ SEQUENCE 602 AA; 63953 MW; 982EF3245D87C43E CRC64;

Query Match 11.2%; Score 87; DB 1; Length 602;  
Best Local Similarity 89.5%; Pred. No. 4.8;  
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 85 ADYDQLVTRVVTHEMAHA 103  
DB 251 ASRYDQLVTRVVTHEMAHA 269

Search completed: August 2, 2004, 14:49:31  
Job time : 6.3 secs

This Page Blank (uspto)

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: August 2, 2004, 14:39:17 ; Search time 29.7 Seconds  
(without alignments)  
1604.150 Million cell updates/sec

Title: US-09-543-407-22

Perfect score: 776

Sequence: 1 MKLLKVAFAAIVVSGSALA.....DSSVMVRQVGFNNATANCY 151

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1017041 seqs, 315518202 residues

Total number of hits satisfying chosen parameters: 1017041

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

SPTREMBL 25:\*

- 1: sp archaea:\*
- 2: sp bacteria:\*
- 3: sp fungi:\*
- 4: sp human:\*
- 5: sp invertebrate:\*
- 6: sp mammal:\*
- 7: sp mhc:\*
- 8: sp organelle:\*
- 9: sp phase:\*
- 10: sp plant:\*
- 11: sp rodent:\*
- 12: sp virus:\*
- 13: sp vertebrate:\*
- 14: sp unclassified:\*
- 15: sp rvirus:\*
- 16: sp bacteriaph:\*
- 17: sp archaeap:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	673	86.7	152	2	Q33802
2	566.5	73.0	150	2	Q7X243
3	538	69.3	149	2	Q7X240
4	496.5	64.0	152	16	Q8CW63
5	421.5	54.3	150	2	Q7X237
6	305.5	39.4	76	2	Q54069
7	122	15.7	29	2	Q9S3J5
8	115	14.8	139	16	Q8EIH3
9	110	14.2	130	16	Q89JI4
10	109.5	14.1	502	16	Q8EJH4
11	109	14.0	1748	5	Q94821
12	109	14.0	2174	16	Q92JU8
13	106.5	13.7	3501	16	Q8Y106
14	106.5	13.7	3552	16	Q8XSD6
15	106	13.7	179	2	Q33801
16	105	13.5	1422	16	Q8EFU3

```

17 104 13.4 714 16 Q7U5X6
18 102 13.1 191 10 Q7XDR3
19 102 13.1 586 5 Q9NAU4
20 98 12.6 614 16 Q7TYR8
21 98 12.6 615 16 P95249
22 96.5 12.4 171 16 Q89JI3
23 96 12.4 2310 5 Q9W2U7
24 95.5 12.3 3659 16 Q98LN6
25 95 12.2 589 16 Q7TX53
26 95 12.2 590 16 Q53309
27 95 12.2 603 16 Q8VU66
28 94.5 12.2 157 16 Q88HG0
29 94.5 12.2 738 5 Q02402
30 94.5 12.2 7716 16 Q7UWZ8
31 94 12.1 407 5 Q20151
32 93.5 12.0 145 16 Q8UGN9
33 93.5 12.0 645 16 Q7UIC5
34 93.5 12.0 646 16 Q53818
35 93.5 12.0 1286 2 Q841Y5
36 93.5 12.0 1333 16 Q8PD38
37 92.5 11.9 151 16 Q7UCZ1
38 92.5 11.9 154 16 Q89JI5
39 92.5 11.9 160 16 Q8CW64
40 92.5 11.9 160 16 Q83PU7
41 92.5 11.9 453 5 Q9NGF7
42 92.5 11.9 453 5 Q9NGM8
43 92 11.9 196 10 Q22638
44 91.5 11.8 151 2 Q7X238
45 91.5 11.8 453 5 Q9NGF6

```

#### ALIGNMENTS

#### RESULT 1

```

O33802 ID O33802 PRELIMINARY; PRT; 152 AA.
AC O33802;
DT 01-JAN-1998 (TREMBLrel. 05, Created)
DT 01-JAN-1998 (TREMBLrel. 05, Last sequence update)
DT 01-DEC-2001 (TREMBLrel. 15, Last annotation update)
DE AG2A protein (Fragment).
GN AG2A.
OS Salmonella typhimurium.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Salmonella.
OX NCBI_TaxID=602;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=98053981; PubMed=9393832;
RA Sukopolvi S.S., Lorentz R.G., Gordon J.I., Bian Z., Pfeifer J.D.,
RA Normark S.J., Rhen M.;
RT "Expression of thin, aggregative fimbriae promotes interaction of
RT Salmonella typhimurium SR-11 with mouse small intestinal epithelial
RT cells."
RL Infect. Immun. 65:5320-5325(1997).
DR EMBL; AJ000514; CA04151.1; -.
FT NON_CODING; 152
SQ SEQUENCE 152 AA; 15401 MW; 9DA7DADC2364B006 CRC64;

```

```

Query Match 86.7%; Score 673; DB 2; Length 152;
Best Local Similarity 89.4%; Pred. No. 3.7e-47;
Matches 135; Conservative 3; Mismatches 13; Indels 0; Gaps 0;
QY 1 MKLLKVAFAAIVVSGSALAGVVPQWGGGNGHNGGNSGSPDSTLSIYQGSANAALALQ 60
DB 1 MKLLKVAFAAIVVSGSALAGVVPQWGGGNGHNGGNSGSPDSTLSIYQGSANAALALQ 60
QY 61 SDARKSETTITQSGYNGADVGQADNYDQLVTRVVTTHMAHADQWNAKNSDITVQYGG 120
DB 61 SDARKSETTITQSGYNGADVGQADNSTIELTQNGFRNNATIDQWNAKNSDITVQYGG 120
QY 121 NNAALVNTASDSSVMVRQVGFNNATANCY 151

```

```

Db      121  NNAALVNQTASDSSVMVRQVGFNNAPANQY 151
|||||
RESULT 2
Q7X243 ID Q7X243 PRELIMINARY; PRT; 150 AA.
AC Q7X243;
DT 01-OCT-2003 (TrEMBLrel. 25, Created)
DT 01-OCT-2003 (TrEMBLrel. 25, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Curlin-csgA protein.
GN CSGA.
OS Citrobacter sp. Fec2.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Citrobacter.
OX NCBI_TaxID=213763;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Fec2;
RA Zogaj X., Bokranz W., Nintz M., Romling U.;
RT "Production of Cellulose and Curli Fimbriae by Members of the Family
RT Enterobacteriaceae Isolated from the Human Gastrointestinal Tract.";
RL Infect. Immun. 72:4151-4158 (2003).
DR EMBL; AJ515700; CAD56672.1;
DR SEQUENCE 150 AA; 15016 MW; 1D7141B8D6973DC6 CRC64;

Query Match 73.0%; Score 566.5; DB 2; Length 150;
Best Local Similarity 77.5%; Pred. No. 1.6e-38;
Matches 117; Conservative 11; Mismatches 22; Indels 1; Gaps 1;

QY 1 MKLLKVAFAAIVVSGSALAGVVPQVGGGNGGNGSGGPDSTLSIYQYGSANAALQ 60
|||||
Db 1 MKLLQVAFAAIVVSGSALAGVVPQVGGGNGGNGSGGPDSTLSIYQYGSANAALQ 59
|||||
QY 61 SDARKSETTITQSGYNGGADVGQADNYDQLVTRVVTHEMAHADOWNAKNSDITVGOYGG 120
|||||
Db 60 SDARKSDTIHQNGFGNGADVGQGSNSIDLTQNGFKNATIDQWNGKNSDITVGOYGG 119
|||||

QY 121 NNAALVNQTASDSSVMVRQVGFNNATANQY 151
|||||
Db 120 HNAALVNQTASDSSVLVHQVGFNNATANQY 150
|||||

RESULT 3
Q7X240 ID Q7X240 PRELIMINARY; PRT; 149 AA.
AC Q7X240;
DT 01-OCT-2003 (TrEMBLrel. 25, Created)
DT 01-OCT-2003 (TrEMBLrel. 25, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Curlin-csgA protein.
GN CSGA.
OS Citrobacter freundii.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Citrobacter.
OX NCBI_TaxID=546;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Fec4;
RA Zogaj X., Bokranz W., Nintz M., Romling U.;
RT "Production of Cellulose and Curli Fimbriae by Members of the Family
RT Enterobacteriaceae Isolated from the Human Gastrointestinal Tract.";
RL Infect. Immun. 72:4151-4158 (2003).
DR EMBL; AJ515701; CAD56675.1;
DR SEQUENCE 149 AA; 15260 MW; 946DD52017F648FD CRC64;

Query Match 69.3%; Score 538; DB 2; Length 149;
Best Local Similarity 72.2%; Pred. No. 3.2e-36;
Matches 109; Conservative 17; Mismatches 23; Indels 2; Gaps 1;

QY 1 MKLLKVAFAAIVVSGSALAGVVPQVGGGNGGNGSGGPDSTLSIYQYGSANAALQ 60
|||||

```

```

Db      1  MKLLKVAFAAIVVSGSALAGVVPQV--GGNHGGGSGNYGPDSSLSIYQYGSNNSANALQ 58
61  SDARKSETTITQSGYNGGADVGQADNYDQLVTRVVTHEMAHADOWNAKNSDITVGOYGG 120
|||||
Db 59  SDARKSDVTITQHGNGGAVVVGQADDSTLSKQTFQNSATIDQWNAKADISVTQFGG 118
|||||
QY 121  NNAALVNQTASDSSVMVRQVGFNNATANQY 151
|||||
Db 119  RNgALVNQTASDSSVLVLIQVGFNNATANQH 149
|||||

RESULT 4
Q8CW63 ID Q8CW63 PRELIMINARY; PRT; 152 AA.
AC Q8CW63;
DT 01-MAR-2003 (TrEMBLrel. 23, Created)
DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE Major curliin subunit precursor.
GN CSGA OR C1306.
OS Escherichia coli O6.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Escherichia.
OX NCBI_TaxID=217992;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=O6:H1 / CFT073 / ATCC 700928;
RX MEDLINE=22388234; PubMed=12471157;
RA Welch R.A., Burland V., Plunkett G. III, Redford P., Roesch P.,
RA Rasko D., Buckles E.L., Liou S.-R., Boutin A., Hackett J., Stroud D.,
RA Mayhew G.F., Rose D.J., Zhou S., Schwartz D.C., Perna N.T.,
RA Mobley H.L.T., Donnenberg M.S., Blattner F.R.;
RT "Extensive mosaic structure revealed by the complete genome sequence
RT of uropathogenic Escherichia coli.";
RL Proc. Natl. Acad. Sci. U.S.A. 99:17020-17024 (2002).
DR EMBL; AE016759; AAN79779.1;
KW Complete proteome.
SQ SEQUENCE 152 AA; 15064 MW; 3BA57F34C1240383 CRC64;

Query Match 64.0%; Score 496.5; DB 16; Length 152;
Best Local Similarity 67.8%; Pred. No. 7.7e-33;
Matches 103; Conservative 17; Mismatches 31; Indels 1; Gaps 1;

QY 1  MKLLKVAFAAIVVSGSALAGVVPQV--GGGNGHGGGNGSGGPDSTLSIYQYGSANAALQ 59
|||||
Db 1  MKLLKVAFAAIVVSGSALAGVVPQVGGGNGGNGSGGNGSGGNSGPNSELNIYQYGGNSALQ 60
|||||
QY 60  QSDARKSETTITQSGYNGGADVGQADNYDQLVTRVVTHEMAHADOWNAKNSDITVGOYGG 119
|||||
Db 61  QADARNSDLTITQHGNGGADVGQGSDDSSIDLTQRFQNSATIDQWNGKSDTMTVQKFG 120
|||||
QY 120  GNNALVNQTASDSSVMVRQVGFNNATANQY 151
|||||
Db 121  GGNGAIVDQTASNSVNVTVQVGFNNATANQY 152
|||||

RESULT 5
Q7X237 ID Q7X237 PRELIMINARY; PRT; 150 AA.
AC Q7X237;
DT 01-OCT-2003 (TrEMBLrel. 25, Created)
DT 01-OCT-2003 (TrEMBLrel. 25, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Curlin-csgA protein.
GN CSGA.
OS Enterobacter sakazakii.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Enterobacter.
OX NCBI_TaxID=28141;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Fec39;
RA Zogaj X., Bokranz W., Nintz M., Romling U.;

```

```

RT "Production of Cellulose and Curli Fimbriae by Members of the Family
RT Enterobacteriaceae Isolated from the Human Gastrointestinal Tract.;"
RL Infect. Immun. 72:4151-4158 (2003).
DR EMBL; AJ515702; CAD56678.1; -.
SQ SEQUENCE 150 AA; 151112 MW; 5D8BB2D872DF15F3 CRC64;

Query Match 54.3%; Score 421.5; DB 2; Length 150;
Best Local Similarity 58.9%; Pred. No. 9.2e-27;
Matches 89; Conservative 25; Mismatches 36; Indels 1; Gaps 1;

QY 1 MKLLKVAFAAIVVSGSALAGVVPQWGGNGGNSGPDSTLSIYQGSANAALALQ 60
Db 1 MKFIKVAALAAIVVSGSAMAGMIIQ-GGWGHGCHGGYGGPNSTLNIYQNGGNSALALQ 59

QY 61 SPARKSETTITQSGYNGADVQCGADNYDQLVTRVVTHEMAHADOWNAKNSDITVGYCG 120
Db 60 TDARNSVLNISOTGGGNGADVQCGSDSSINFQNGFGNSATLDQNSKDSVMVNSYQYG 119

QY 121 NNAALVNQTASDSSVMVROVFGNNATANQY 151
Db 120 LAGALVDQTASNSTVNVTQIFGNHATAHQY 150

RESULT 6
Q54069 PRELIMINARY; PRT; 76 AA.
AC Q54069;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE SEF17 fimbria (Fragment).
GN AGFA.
OS Salmonella enteritidis.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Salmonella.
OX NCBI_TaxID=592;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=SE30;
RA Cox J.M.; Egelezos S.; Woolcock J.B.;
RT "virulence of Salmonella enteritidis in chickens correlates with
RT colony morphology and expression of SEF17 fimbriae";
RL Submitted (APR-1996) to the EMBL/GenBank/DDBJ databases.
DR EMBL; U53207; AAA98671.1; -.
FT NON-TER 1
FT NON-TER 76
SQ SEQUENCE 76 AA; 7704 MW; 2PD5411241A7BCB1 CRC64;

Query Match 39.4%; Score 305.5; DB 2; Length 76;
Best Local Similarity 54.5%; Pred. No. 1e-17;
Matches 66; Conservative 4; Mismatches 6; Indels 45; Gaps 1;

QY 30 GNGHGGNSGPDSTLSIYQGSANAALALQSDARKSETTITQSGYNGADVQCGADNYD 89
Db 1 GNGXGGNSGPDSTLSIYQGSANAALALQSDARKSETTITQSGYNGADVQCGADN-- 58

QY 90 QLVTRVVTHEMAHADOWNAKNSDITVGYCGNNAALVNQTASDSSVMVROVFGNNATAN 149
Db 59 -----STIELTQNGFRNNATID 75

QY 150 Q 150
Db 76 Q 76

RESULT 7
Q9S3J5 PRELIMINARY; PRT; 29 AA.
AC Q9S3J5;
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT 01-MAR-2001 (TrEMBLrel. 16, Last annotation update)
DE Curlin subunit monomer (Fragment).

```

```

CGSA.
OS Escherichia coli.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Escherichia.
OX NCBI_TaxID=562;
RN [1]
RP SEQUENCE FROM N.A.
RC TRANSPOSON=Insertion sequence IS1;
RC MEDLINE=99314153; PubMed=10386375;
LA Regione R.M., Collighan R.J., Woodward M.J.;
RT "Non-curliation of Escherichia coli O78:K80 isolates associated with
RT IS1 inserti on in csGB and reduced persistence in poultry infection.;"
RL FEMS Microbiol. Lett. 175:247-253 (1999).
DR EMBL; AJ131756; CAB45380.1; -.
FT NON-TER 29
FT NON-TER 29
SQ SEQUENCE 29 AA; 2789 MW; E290DFC07ABB243 CRC64;

Query Match 15.7%; Score 122; DB 2; Length 29;
Best Local Similarity 89.7%; Pred. No. 0.0025;
Matches 26; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 MKLLKVAFAAIVVSGSALAGVVPQWGGG 29
Db 1 MKLLKVAALAAIVFSGSALAGVVPQYGGG 29

RESULT 8
Q8EIH3 PRELIMINARY; PRT; 139 AA.
AC Q8EIH3;
DT 01-MAR-2003 (TrEMBLrel. 23, Created)
DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE Minor curlin subunit CsgB, putative.
GN SO0866.
OS Shewanella oneidensis.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Alteromonadales;
OC Alteromonadaceae; Shewanella.
OX NCBI_TaxID=70863;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=MR-1;
RC MEDLINE=22297686; PubMed=12368813;
RA Heidelberg J.F.; Paulsen I.T.; Nelson K.E.; Gaidos E.J.; Nelson W.C.;
RA Read I.D.; Eisen J.A.; Seshadri R.; Ward N.; Methe B.; Clayton R.A.;
RA Meyer T.; Tsapin A.; Scott J.; Beanan M.; Brinkac L.; Daugherty S.;
RA DeBoy R.T.; Dodson R.J.; Durkin A.S.; Haft D.H.; Kolonay J.F.;
RA Madupu R.; Peterson J.D.; Umayam L.A.; White O.; Wolf A.M.;
RA Vamathevan J.; Weidman J.; Imbraim M.; Lee K.; Berry K.; Lee C.;
RA Mueller J.; Khouri H.; Gill J.; Utterback T.R.; McDonald L.A.;
RA Feldblyum T.V.; Smith H.O.; Venter J.C.; Nealon K.H.; Fraser C.M.;
RT "Genome sequence of the dissimilatory metal ion-reducing bacterium
RT Shewanella oneidensis.;"
RL Nat. Biotechnol. 20:1118-1123 (2002).
DR EMBL; AE015532; AAN53942.1; -.
DR TIGR; SO0866; -.
KW Complete proteome.
SQ SEQUENCE 139 AA; 14811 MW; 41EC1CFA76957920 CRC64;

Query Match 14.8%; Score 115; DB 16; Length 139;
Best Local Similarity 28.3%; Pred. No. 0.061;
Matches 30; Conservative 19; Mismatches 33; Indels 24; Gaps 2;

QY 39 SGPDSLTLSIYQGSANAALALQSDARKSETTITQSGYNGADVQCGADNYDQLVTRVVT 98
Db 41 SGDRNLIDLIVQQTANQGI-----VFQSGSDNSAVVTQAGNDNISLVTQIGT- 87

QY 99 EMAHADQWNAKNSDITVGYCGNNAALVNQTASDSSVMVROVFGN 144
Db 88 -----NNEVQLLVQGAQNAKASITQIGNDNLVQLNLQSGN 122

RESULT 9

```

Q89JI4  
ID Q89JI4 PRELIMINARY; PRT; 130 AA.  
AC Q89JI4;  
DT 01-JUN-2003 (TrEMBLrel. 24, Created)  
DT 01-JUN-2003 (TrEMBLrel. 24, Last sequence update)  
DE 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)  
DE BIL5299 protein.  
GN BIL5299  
OS Bradyrhizobium japonicum.  
OC Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;  
OC Bradyrhizobiaceae; Bradyrhizobium.  
OX NCBI\_TaxID=375;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=USDA 110;  
RX MEDLINE=22484998; PubMed=12597275;  
RA Kaneko T., Nakamura Y., Sato S., Minamisawa K., Uchiumi T.,  
RA Sasamoto S., Watanabe A., Idesawa K., Iriiguchi M., Kawashima K.,  
RA Kohara M., Matsumoto M., Shimpo S., Tsuruoka H., Wada T., Yamada M.,  
RA Tabata S.;  
RT "Complete genomic sequence of nitrogen-fixing symbiotic bacterium  
Bradyrhizobium japonicum USDA110.";  
RL DNA Res. 9:189-197(2002).  
DR EMBL; AP005954; BAC50564.1; -.  
KW Complete proteome.  
SQ SEQUENCE 130 AA; 12699 MW; ACFB2D66A48D260F CRC64;  
  
Query Match 14.2%; Score 110; DB 16; Length 130;  
Best Local Similarity 25.8%; Pred. No. 0.14; Mismatches 28; Indels 28; Gaps 5;  
Matches 39; Conservative  
  
QY 4 LKVAFAAIVVGSALAGVVPQGGNGGNGGNSGGPSTLSIYQGSANAALALQSDA 63  
DB 1 MRLTYLVATAIALSALTUTVDAQ-----AGNSA-----SVLQFGTNSFSIQGS 45  
  
QY 64 RKSETITQSGYGVGADVGCGADNYDQVTRVVTHEMAHADW-----NAKNSDITVGVG 119  
DB 46 TSNNAITLQFGATNTATTIQTGS-----LLVTNTAVTGQGTATASNTALT-GQVG 96  
  
QY 120 GNNAAALVNQTSDDSSVMYRVQVFGNNATANQ 150  
DB 97 GSNSSLIQIGANNTAGVQGLGILNGSTILQ 127  
  
RESULT 10  
ID Q8EIH4 PRELIMINARY; PRT; 502 AA.  
AC Q8EIH4;  
DT 01-MAR-2003 (TrEMBLrel. 23, Created)  
DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)  
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)  
DE Conserved hypothetical protein.  
GN S00865.  
OS Shewanella oneidensis.  
OC Bacteria; Proteobacteria; Gammaproteobacteria; Alteromonadales;  
OC Alteromonadaceae; Shewanella.  
OX NCBI\_TaxID=70863;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=MR-1;  
RX MEDLINE=2297686; PubMed=12368813;  
RA Heidelberg J.F., Paulsen I.T., Nelson K.E., Gaidos E.J., Nelson W.C.,  
RA Read T.D., Eisen J.A., Seshadri R., Ward N., Methe B., Clayton R.A.,  
RA Meyer T., Tsapin A., Scott J., Beanan M., Brinkac L., Daugherty S.,  
RA DeBoy R.T., Dodson R.J., Durkin A.S., Haft D.H., Kolonay J.F.,  
RA Madupu R., Peterson J.D., Unayam L.A., White O., Wolf A.M.,  
RA Vamathevan J., Weidman J., Impriali M., Lee K., Berry K., Lee C.,  
RA Mueller J., Khouri H., Gill J., Utterback T.R., McDonald L.A.,  
RA Feldblyum T.V., Smith H.O., Venter J.C., Nealon K.H., Fraser C.M.;  
RT "Genome sequence of the dissimilatory metal ion-reducing bacterium  
Shewanella oneidensis.";  
RL Nat. Biotechnol. 20:1118-1123(2002).  
DR EMBL; AB015532; AAN53941.1; -.

TIGR; S00865; -.  
KW Hypothetical protein; Complete proteome.  
SQ SEQUENCE 502 AA; 52441 MW; D08CA23D6C46B62D CRC64;  
  
Query Match 14.1%; Score 109.5; DB 16; Length 502;  
Best Local Similarity 23.8%; Pred. No. 0.79;  
Matches 39; Conservative  
  
QY 29 GGNHNGGNSGGPSTLSIYQGSANAA-----LALQSDARKSETTITQSGVGG 78  
DB 347 GDNELVAFATGEDNSIEISQEGDANPAYDATGNNEVNVQGGQNETIIVG-NNN 405  
QY 79 ADVG-----QCADNYDQVTRVVTHEMAHADW-----N 107  
DB 406 ADVTALQHRGDLNLIDLIEGDENAAEIT-----QAGSGNVWGGDSSSSFAASFGVS 458  
QY 108 AKNSDITVGVGNNAAALVNQTSDDSSVMYRVQVFGNNATANQ 151  
DB 459 GDNNSLMITGTGNDNLVLGSAQGNNSISVTQSGDMNVATVQY 502  
  
RESULT 11  
ID Q94821 PRELIMINARY; PRT; 1748 AA.  
AC Q94821; P92146; P92145; P92144; P92143; P92142; P92141; Q94820;  
DT 01-FEB-1997 (TrEMBLrel. 02, Created)  
DT 01-FEB-1997 (TrEMBLrel. 02, Last sequence update)  
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)  
DE CNJB protein.  
GN CNJB  
OS Tetrahymena thermophila.  
OC Eukaryota; Alveolata; Ciliophora; Oligohymenophorea; Hymenostomatida;  
OC Tetrahymenina; Tetrahymena.  
OX NCBI\_TaxID=5911;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=98189811; PubMed=3357771;  
RA Martindale D.W., Taylor F.M.;  
RT "Multiple introns in a conjugation-specific gene from Tetrahymena  
thermophila.";  
RL Nucleic Acids Res. 16:2189-2201(1988).  
RN [2]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=94051569; PubMed=8233798;  
RA Taylor F.M., Martindale D.W.;  
RT "Retroviral-type zinc fingers and glycine-rich repeats in a protein  
encoded by cnjB, a Tetrahymena gene active during meiosis.";  
RL Nucleic Acids Res. 21:4610-4614(1993).  
DR EMBL; X06462; CAB37323.1; -.  
DR EMBL; L03710; AAC37171.1; -.  
DR PIR; S42136; S42136.  
DR HSSP; P05888; IAAF.  
DR GO; GO:0003676; F:nucleic acid binding; IEA.  
DR InterPro; IPR001878; Znf\_CCHC.  
DR Pfam; PF00098; zf-CCHC; 7.  
DR PRINTS; PR00939; C2HCZNFINGER.  
DR SMART; SMO0343; Znf\_C2HC; 7.  
DR PROSITE; PS00159; ZF\_CCHC; 7.  
DR CONFLICT 251 251 M -> I (IN REF. 1).  
DR CONFLICT 256 256 I -> N (IN REF. 1).  
SQ SEQUENCE 1748 AA; 199624 MW; 0803F210104008A3 CRC64;  
  
Query Match 14.0%; Score 109; DB 5; Length 1748;  
Best Local Similarity 33.6%; Pred. No. 3.9;  
Matches 37; Conservative  
  
QY 25 QWGGGNGHNGG---GNSSGPDSTLSIYQGSANAALALQSDARKSETTIT---QSGYGN 77  
DB 1640 QFGGGGNGGNGGQSGWTSSGSDWN-----CQSNVOESTTTSGGGWSGSGN 1685  
  
QY 78 GADYVQGA-DNYDQVTRVVTHEMAHADWNAKNSDITVGVG---YGGNAA 124  
DB 1686 QTGGGWSNDNQOQ-----QNETGGGGWSNSQTNNESSWGSNQA 1729



```

RESULT 12
Q92U08 PRELIMINARY; PRT; 2174 AA.
AC Q92U08;
DT 01-DEC-2001 (TrEMBLrel. 19, Created)
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Hypothetical glycine-rich protein SM21548.
GN R80989 OR SM21548.
OS Rhizobium meliloti (Sinorhizobium meliloti).
OG Plasmid pSymB (megaplasmid 2).
OC Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;
OC Rhizobiaceae; Sinorhizobium/Ensifer group; Sinorhizobium.
OX NCBI_TaxID=382;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=1021;
RX MEDLINE=21396508; PubMed=11481431;
RA Finan T.M., Weidner S., Wong K., Buhrmester J., Chain P.,
RA Vorholter F.J., Hernandez-Lucas I., Becker A., Cowie A., Gouzy J.,
RA Golding B., Puehler A.;
RT "The complete sequence of the 1,693-kb pSymB megaplasmid from the N2-
fixing endosymbiont Sinorhizobium meliloti."
RL Proc. Natl. Acad. Sci. U.S.A. 98:9889-9894(2001).
DR EMBL; AL603645; CAC49389.1; -.
DR PIR; E95965; E95965.
DR GO; GO:0046821; C:extrachromosomal DNA; IEA.
DR GO; GO:0004601; F:peroxidase activity; IEA.
DR GO; GO:0006979; P:response to oxidative stress; IEA.
DR InterPro; IPR005546; Autotransporter.
DR InterPro; IPR002016; Peroxidase.
DR InterPro; IPR002173; Pfkb.
DR Pfam; PF03797; Autotransporter; 1.
DR PROSITE; PS00435; PEROXIDASE_1; 1.
DR PROSITE; PS00583; PFKE_KINASES_1; 2.
DR Plasmid; Hypothetical protein; Complete proteome.
KW SEQUENCE 2174 AA; 203314 MW; 008EB68297B44182 CRC64;
SQ SEQUENCE 2174 AA; 203314 MW; 008EB68297B44182 CRC64;

Query Match 14.0%; Score 109; DB 16; Length 2174;
Best Local Similarity 27.0%; Pred. No. 5.1;
Matches 40; Conservative 20; Mismatches 52; Indels 36; Gaps 7;

QY 11 AVVSGSALAGVVPQ--WGCGNHNGGNSGPDSTLSIYQVGS-----ANAA----- 56
DB 693 AIATAGAGAVGILAQSIGGGGNGN---GGNATGAGAGFGSGGGGGGGYANTANVGFK 749
QY 57 -LALQSDARKSETTITQSGYGNAGDYVGQADNYDQLVTRVWTHMAHQDQWNAKNSDITV 115
DB 750 GLTLTQGGHAGIVAQSS-VGGGGGTGCTASSYSAGI-----GFTASVAV 793
QY 116 GQYGGNAA--LVNOTASDSSVMVRQV 141
DB 794 GGTGGNGGAGGEVSVSLTDSAIRTGQGG 821

RESULT 13
Q8Y106 PRELIMINARY; PRT; 3501 AA.
AC Q8Y106;
DT 01-MAR-2002 (TrEMBLrel. 20, Created)
DT 01-MAR-2002 (TrEMBLrel. 20, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Probable hemagglutinin-related protein.
GN RSC0887 OR RSC06116.
OS Ralstonia solanacearum (Pseudomonas solanacearum).
OC Bacteria; Proteobacteria; Betaproteobacteria; Burkholderiales;
OC Burkholderiaceae; Ralstonia.
OX NCBI_TaxID=305;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=GM11000;
RX MEDLINE=21681879; PubMed=11823852;
RA Salanoubat M., Genin S., Artiguenave F., Gouzy J., Mangenot S.,
RA Arlat M., Billault A., Brottier P., Camus J.C., Cattolico L.,
RA Chandler M., Choisme N., Claudel-Renard C., Cunnac S., Demange N.,
RA Gaspin C., Lavie M., Moisan A., Robert C., Saurin W., Schiex T.,
RA Sigulier P., Thebault P., Whalen M., Wincker P., Levy M.,
RA Weissenbach J., Boucher C.A.;
RT "Genome sequence of the plant pathogen Ralstonia solanacearum."
RL Nature 415:497-502(2002).
DR EMBL; AL646061; CAD14589.1; -.
DR GO; GO:0004519; F:endonuclease activity; IEA.
DR GO; GO:0003676; F:nucleic acid binding; IEA.
DR InterPro; IPR001604; Endonuclease.
DR InterPro; IPR008619; Fil_haemagg.
DR Pfam; PF05594; Fil_haemagg; 20.
DR PROSITE; PS01070; NUCLEASE_NON_SPEC; 1.
DR Complete proteome.
KW SEQUENCE 3501 AA; 348421 MW; 290B41C99018A107 CRC64;
SQ SEQUENCE 3501 AA; 348421 MW; 290B41C99018A107 CRC64;

Query Match 13.7%; Score 106.5; DB 16; Length 3501;
Best Local Similarity 29.5%; Pred. No. 14;
Matches 43; Conservative 19; Mismatches 45; Indels 39; Gaps 7;

QY 15 SGSALAGVVPQWGCGNHNGG--NSSGPDSTLSIYQVGSANAA----- 56
DB 2431 SGSHFSTAGPSWGLGRNVGGPNSSG---VGLAPYGSASHADNAAGNSRQNASVVG 2486
QY 57 LALQSDARKSETTITQSGYGNAGDYVG---QGADNYDQLVTRVWTHMA--HADQWNAK 110
DB 2487 KSVQVQARTGDIIVSGSGISALSDVLLAKQK-----VDIVAGNDSRSHED-----H 2535
QY 111 SDITVQYGGNNAALVNQTSADSSVM 136
DB 2536 SDRTIGDLGGNGYSGTVGVSASSTL 2561

RESULT 14
Q8XSD6 PRELIMINARY; PRT; 3552 AA.
AC Q8XSD6;
DT 01-MAR-2002 (TrEMBLrel. 20, Created)
DT 01-MAR-2002 (TrEMBLrel. 20, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Probable hemagglutinin-related protein.
GN RSP0540 OR RSP06117.
OS Ralstonia solanacearum (Pseudomonas solanacearum).
OC Bacteria; Proteobacteria; Betaproteobacteria; Burkholderiales;
OC Burkholderiaceae; Ralstonia.
OX NCBI_TaxID=305;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=GM11000;
RX MEDLINE=21681879; PubMed=11823852;
RA Salanoubat M., Genin S., Artiguenave F., Gouzy J., Mangenot S.,
RA Arlat M., Billault A., Brottier P., Camus J.C., Cattolico L.,
RA Chandler M., Choisme N., Claudel-Renard C., Cunnac S., Demange N.,
RA Gaspin C., Lavie M., Moisan A., Robert C., Saurin W., Schiex T.,
RA Sigulier P., Thebault P., Whalen M., Wincker P., Levy M.,
RA Weissenbach J., Boucher C.A.;
RT "Genome sequence of the plant pathogen Ralstonia solanacearum."
RL Nature 415:497-502(2002).
DR EMBL; AL646079; CAD17691.1; -.
DR GO; GO:0046821; C:extrachromosomal DNA; IEA.
DR GO; GO:0004519; F:endonuclease activity; IEA.
DR GO; GO:0003676; F:nucleic acid binding; IEA.
DR InterPro; IPR001604; Endonuclease.
DR InterPro; IPR008619; Fil_haemagg.
DR Pfam; PF05594; Fil_haemagg; 20.
DR PROSITE; PS01070; NUCLEASE_NON_SPEC; 1.
DR Complete proteome.
KW SEQUENCE 3552 AA; 348421 MW; 290B41C99018A107 CRC64;
SQ SEQUENCE 3552 AA; 348421 MW; 290B41C99018A107 CRC64;

```

DR PROSITE; PS01070; NUCLEASE\_NON\_SPEC; 1.  
KW Plasmid; Complete proteome.  
SQ SEQUENCE 3552 AA; 352934 MW; C5432AABE2CCF59C CRC64;

Query Match 13.7%; Score 106.5; DB 16; Length 3552;  
Best Local Similarity 29.5%; Pred. No. 15;  
Matches 43; Conservative 19; Mismatches 45; Indels 39; Gaps 7;

QY 15 GSAAGVVPQGGGNGHNGG-NSSGPDSTLSIYQGSANAA-----56  
Db 2430 SGSHFSTAGSMDLGRNVGGGPNSSG---VGLAPYGSADNAAGNSRQNASVVIG 2485  
QY 57 LALQSDARKSETITQSGYNGADYV---QGADNYDQLVTRVVTHEMA--HADQWNAKN 110  
Db 2486 KSVQVQARTGDIIVSGSGISALSDVLLAKQK-----VDIVAGNDYSSRHE-----H 2534

QY 111 SDITVQYGGNNAALVNQTASDSSVM 136  
Db 2535 SDRTIGLGGNGYSGTVGVSASSTL 2560

RESULT 15  
C033801 PRELIMINARY; PRT; 179 AA.  
AC C033801;  
DT 01-JAN-1998 (TREMBLrel. 05, Created)  
DT 01-JAN-1998 (TREMBLrel. 05, Last sequence update)  
DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)  
DE AgfB protein.  
GN AGFB.  
OS Salmonella typhimurium.  
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;  
OC Enterobacteriaceae; Salmonella.  
OX NCBI\_TaxID=602;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=98053981; PubMed=9393832;  
RA Sukupolvi S.S., Lorentz R.G., Gordon J.I., Bian Z., Pfeifer J.D.,  
RA Normark S.J., Rhen M.;  
RT "Expression of thin, aggregative fimbriae promotes interaction of  
RT Salmonella typhimurium SR-11 with mouse small intestinal epithelial  
RT cells.";  
RL Infect. Immun. 65:5320-5325(1997).  
DR EMBL; AJ000514; CAA04150.1; -.  
SQ SEQUENCE 179 AA; 19318 MW; A2BCCB648B3C0B0B CRC64;

Query Match 13.7%; Score 106; DB 2; Length 179;  
Best Local Similarity 32.3%; Pred. No. 0.44;  
Matches 41; Conservative 17; Mismatches 55; Indels 14; Gaps 5;

QY 16 GSAAGVVPQGGGNGHNGGNSGPDSTLSIYQGSANAAALQS---DARKSETTITQ 72  
Db 63 GSKLLSVISQ-ERGGNRAKVDQAGNPFAYIEQTGNANDASISQAYGNSAAISAAITQ 121  
QY 73 SGYNGADYVQGGADNYDQLVTRVVTHEMAHADQWNAKNSDITVQYGGNNAALVNQTASD 132  
Db 122 KSGGNKANITQ-----YGTOKTAVVQKQSHM-----AIQANIT--QYGTQKTAVVQKQSH 171

QY 133 SSVWVRQ 139  
Db 172 MAIRVTQ 178

Search completed: August 2, 2004, 14:54:40  
Job time : 30.7 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: August 2, 2004, 14:35:42 ; Search time 44.9 Seconds

(without alignments)

950.215 Million cell updates/sec

Title: US-09-543-407-24

Perfect score: 775

Sequence: 1 MKLLKVAFAAIVVSGSALA.....DSSVMVRQVGFNNATANQY 151

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1586107 seqs, 282547505 residues

Total number of hits satisfying chosen parameters: 1586107

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : A\_Geneseq\_25Jan04:\*

1: Genesep1980s:\*

2: Genesep1990s:\*

3: Genesep2000s:\*

4: Genesep2001s:\*

5: Genesep2002s:\*

6: Genesep2003as:\*

7: Genesep2003bs:\*

8: Genesep2004s:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	775	100.0	151	3 AAB36352	Aab36352 Agfa::PT3
2	714	92.1	151	3 AAB36347	Aab36347 Agfa::PT3
3	700	90.3	151	3 AAB36355	Aab36355 Agfa::PT3
4	690	89.0	151	2 AAR74625	Aar74625 Agfa sequ
5	690	89.0	151	3 AAB36341	Aab36341 Salmonell
6	689	88.9	151	2 AAW23570	Aaw23570 Salmonell
7	655	84.5	151	3 AAB36346	Aab36346 Agfa::PT3
8	609	78.6	151	3 AAB36353	Aab36353 Agfa::PT3
9	609	78.6	151	3 AAB36351	Aab36351 Agfa::PT3
10	608	78.5	151	3 AAB36349	Aab36349 Agfa::PT3
11	606	78.2	151	3 AAB36350	Aab36350 Agfa::PT3
12	600	77.4	151	3 AAB36354	Aab36354 Agfa::PT3
13	575	74.2	151	3 AAB36348	Aab36348 Agfa::PT3
14	533	68.8	151	3 AAB36343	Aab36343 Escherich
15	528	68.1	151	7 ABR82651	Abr82651 E. coli C
16	505	65.2	120	2 AAR62761	Aar62761 Agfa sequ
17	505	65.2	120	2 AAW23569	Aaw23569 Salmonell
18	455	58.7	142	2 AAR52664	Aar52664 Fibronect
19	383	49.4	122	2 AAR52663	Aar52663 FNB curli
20	237	30.6	45	3 AAB36316	Aab36316 Salmonell
21	132	17.0	22	3 AAB36318	Aab36318 Salmonell
22	123	15.9	23	3 AAB36321	Aab36321 Salmonell
23	123	15.9	23	3 AAB36326	Aab36326 Salmonell
24	123	15.9	23	3 AAB36338	Aab36338 Salmonell
25	115	14.8	22	3 AAB36325	Aab36325 Salmonell

26	115	14.8	22	3 AAB36339	Aab36339 Salmonell
27	115	14.8	22	3 AAB36320	Aab36320 Salmonell
28	113	14.6	24	7 ABR82644	Abr82644 E. coli C
29	109	14.1	23	3 AAB36340	Aab36340 Salmonell
30	109	14.1	23	3 AAB36324	Aab36324 Salmonell
31	109	14.1	23	3 AAB36319	Aab36319 Salmonell
32	98.5	12.7	151	3 AAB36342	Aab36342 Salmonell
33	98	12.6	26	7 ABR82649	Abr82649 E. coli V
34	98	12.6	26	7 ABR82645	Abr82645 E. coli C
35	98	12.6	903	2 AAW35006	Aaw35006 Polyangi
36	96.5	12.5	151	3 AAB36344	Aab36344 Escherich
37	96	12.4	19	3 AAB36323	Aab36323 Salmonell
38	96	12.4	19	3 AAB36336	Aab36336 Salmonell
39	96	12.4	19	3 AAB36328	Aab36328 Salmonell
40	92.5	11.9	850	4 ABB65764	Abb65764 Drosophil
41	92.5	11.9	1028	4 ABB62708	Abb62708 Drosophil
42	92	11.9	23	3 AAB36331	Aab36331 Escherich
43	91	11.7	688	5 ABB74039	Abb74039 Candida a
44	90.5	11.7	378	4 ABB66461	Abb66461 Drosophil
45	90.5	11.7	520	6 AAO16497	Aao16497 Argiope t

ALIGNMENTS

RESULT 1  
AAB36352  
ID AAB36352 standard; protein; 151 AA.  
XX AAB36352;  
AC AAB36352;  
XX 26-FEB-2001 (first entry)  
XX  
DE Agfa::PT3#7 amino acid sequence SEQ ID NO:24.  
XX  
KW Salmonella; agfa; chromosomal gene replacement; fimbria; epitope;  
KW vaccine; immune response; immunogen.  
XX  
OS Salmonella enteritidis.  
OS Escherichia coli.  
OS Synthetic.  
XX  
PN WO2000060102-A2.  
XX  
PD 12-OCT-2000.  
XX  
PF 05-APR-2000; 2000WO-CA000356.  
XX  
PR 05-APR-1999; 99US-0127888P.  
XX  
PA (UYVI-) UNIV VICTORIA.  
XX  
PI White AP, Doran JL, Collison SK, Kay WW;  
XX  
DR WPI; 2000-672631/65.  
XX  
DR N-PSDB; AAC64628.  
XX  
PT Recombinant agfa gene having a segment replaced by a foreign DNA sequence  
PT which encodes foreign epitope or antigen, expresses recombinant Agfa  
PT protein useful for eliciting immune response in animal.  
XX  
PS Disclosure; Page 138; 139pp; English.  
XX  
CC The present invention describes a recombinant agfa gene (I) where a  
CC segment of the gene has been replaced by a segment of a foreign DNA  
CC sequence which encodes a foreign epitope or antigen. Also described are:  
CC (1) use of thin aggregative fimbriae (SEF17/TAF) nucleation depended  
CC assembly system of strains of Salmonella, Escherichia coli and  
CC Enterobacteriaceae for the production of fimbriae comprising recombinant  
CC Agfa, CsgA and agfa-homologue fimbria subunits, respectively; (2)  
CC directing recombination of a recombinant gene into the chromosome of the  
CC homologous species; (3) directing recombination of a recombinant gene  
CC back into the chromosome of the homologous species, replacing the native

CC copy of that gene; and (4) eliciting an immune response in an animal,  
CC comprising separating an amino acid polymer comprising a recombinant Agfa  
CC protein containing a replacement segment or segments of foreign amino  
CC acid sequence or sequences grown on a Salmonella, E. coli or  
CC Enterobacteriaceae host cell, from the host cell and introducing the  
CC polymer into the animal in conjunction with a carrier or diluent. (1) is  
CC useful for the expression of recombinant Agfa protein which is useful for  
CC eliciting an immune response in an animal. In a fibrial presentation  
CC system the heterologous antigens are presented in high numbers (up to  
CC 500,000 copies/cell), the hybrid fimbrial protein possesses both the  
CC immunogenicity and adhesion properties relevant for an efficient live  
CC vaccine, the carrier fimbrial subunit proteins are usually strong  
CC immunogens, which may be important for directing an immune response  
CC against the inserted epitope, and hybrid fimbriae are easy and  
CC inexpensive to purify in large amount. The present sequence is given in  
CC the exemplification of the present invention  
XX  
XX Sequence 151 AA;

Query Match 100.0%; Score 775; DB 3; Length 151;  
Best Local Similarity 100.0%; Pred. No. 1.6e-68;  
Matches 151; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 MKLLKVAFAAIVVSGSALAGVVPQWGGGNGHNGSGSPDSTLSIYQGSANAALALQ 60  
DB 1 MKLLKVAFAAIVVSGSALAGVVPQWGGGNGHNGSGSPDSTLSIYQGSANAALALQ 60  
QY 61 SDARKSETTITQSGYNGADVGQADNSTIELTQNGFRNNATIDQWNAKNDQLVTRVVT 120  
DB 61 SDARKSETTITQSGYNGADVGQADNSTIELTQNGFRNNATIDQWNAKNDQLVTRVVT 120  
QY 121 HEWAHANQATASDSSVMVRQVGFNNATANQY 151  
DB 121 HEWAHANQATASDSSVMVRQVGFNNATANQY 151

RESULT 2  
AAB36347  
ID AAB36347 standard; protein; 151 AA.  
AC AAB36347;  
DT 26-FEB-2001 (first entry)  
XX Agfa::PT3#2 amino acid sequence SEQ ID NO:14.  
DE Salmonella; agfa; chromosomal gene replacement; fimbria; epitope;  
KW vaccine; immune response; immunogen.  
XX Salmonella enteritidis.  
OS Escherichia coli.  
OS Synthetic.  
XX WO2000060102-A2.  
PN 12-OCT-2000.  
PD 05-APR-2000; 2000WO-CA000356.  
PF 05-APR-1999; 99US-0127888P.  
XX (UYVI-) UNIV VICTORIA.  
PA White AP, Doran JL, Collison SK, Kay WW;  
FI WPI; 2000-672631/65.  
DR N-PSDB; AAC64623.  
XX Recombinant agfa gene having a segment replaced by a foreign DNA sequence  
PT which encodes foreign epitope or antigen, expresses recombinant Agfa  
PT protein useful for eliciting immune response in animal.  
XX Disclosure; Page 136; 139pp; English.  
PS

XX The present invention describes a recombinant agfa gene (1) where a  
CC segment of the gene has been replaced by a segment of a foreign DNA  
CC sequence which encodes a foreign epitope or antigen. Also described are:  
CC (1) use of thin aggregative fimbriae (SEF17/7AF) nucleation depended  
CC assembly system of strains of Salmonella, Escherichia coli and  
CC Enterobacteriaceae for the production of fimbriae comprising recombinant  
CC Agfa, CsgA and Agfa-homologue fimbrial subunits, respectively; (2)  
CC directing recombination of a recombinant gene into the chromosome of the  
CC homologous species; (3) directing recombination of a recombinant gene  
CC back into the chromosome of the homologous species, replacing the native  
CC copy of that gene; and (4) eliciting an immune response in an animal,  
CC comprising separating an amino acid polymer comprising a recombinant Agfa  
CC protein containing a replacement segment or segments of foreign amino  
CC acid sequence or sequences grown on a Salmonella, E. coli or  
CC Enterobacteriaceae host cell, from the host cell and introducing the  
CC polymer into the animal in conjunction with a carrier or diluent. (1) is  
CC useful for the expression of recombinant Agfa protein which is useful for  
CC eliciting an immune response in an animal. In a fibrial presentation  
CC system the heterologous antigens are presented in high numbers (up to  
CC 500,000 copies/cell), the hybrid fimbrial protein possesses both the  
CC immunogenicity and adhesion properties relevant for an efficient live  
CC vaccine, the carrier fimbrial subunit proteins are usually strong  
CC immunogens, which may be important for directing an immune response  
CC against the inserted epitope, and hybrid fimbriae are easy and  
CC inexpensive to purify in large amount. The present sequence is given in  
CC the exemplification of the present invention  
XX  
XX Sequence 151 AA;

Query Match 92.1%; Score 714; DB 3; Length 151;  
Best Local Similarity 91.1%; Pred. No. 1.8e-62;  
Matches 144; Conservative 0; Mismatches 0; Indels 14; Gaps 2;  
QY 1 MKLLKVAFAAIVVSGSALAGVVPQWGGGNGHNGSGSPDSTLSIYQGSANAALALQ 60  
DB 1 MKLLKVAFAAIVVSGSALAGVVPQWGGGNGHNGSGSPDSTLSIYQGSANAALALQ 60  
QY 61 SDARKSETTITQSGYNGADVGQADNSTIELTQNGFRNNATIDQWNAKNDQLVTRVVT 113  
DB 61 SDARKSETTITQSGYNGADVGQADNSTIELTQNGFRNNATIDQWNAKNDQLVTRVVT 120  
QY 114 LVTRVVTHEMAHANQATASDSSVMVRQVGFNNATANQY 151  
DB 121 LVTRVVTHEMAHANQATASDSSVMVRQVGFNNATANQY 151

RESULT 3  
AAB36355  
ID AAB36355 standard; protein; 151 AA.  
AC AAB36355;  
XX 26-FEB-2001 (first entry)  
DT Agfa::PT3#10 amino acid sequence SEQ ID NO:30.  
XX Salmonella; agfa; chromosomal gene replacement; fimbria; epitope;  
KW vaccine; immune response; immunogen.  
XX Salmonella enteritidis.  
OS Escherichia coli.  
OS Synthetic.  
XX WO2000060102-A2.  
PN 12-OCT-2000.  
PD 05-APR-2000; 2000WO-CA000356.  
PF 05-APR-1999; 99US-0127888P.  
XX (UYVI-) UNIV VICTORIA.  
PA

XX White AP, Doran JL, Collison SK, Kay WW;  
 XX WPI: 2000-672631/65.  
 DR N-PSDB; AAC64631.  
 XX Recombinant agfa gene having a segment replaced by a foreign DNA sequence  
 PT which encodes foreign epitope or antigen, expresses recombinant Agfa  
 PT protein useful for eliciting immune response in animal.  
 XX Disclosure; Page 139; 139pp; English.  
 XX The present invention describes a recombinant agfa gene (I) where a  
 CC segment of the gene has been replaced by a segment of a foreign DNA  
 CC sequence which encodes a foreign epitope or antigen. Also described are:  
 CC (1) use of thin aggregative fimbriae (SEPI7/TAF) nucleation depended  
 CC assembly system of strains of Salmonella, Escherichia coli and  
 CC Enterobacteriaceae for the production of fimbriae comprising recombinant  
 CC Agfa, CsgA and Agfa-homologue fimbriin subunits, respectively; (2)  
 CC directing recombination of a recombinant gene into the chromosome of the  
 CC homologous species; (3) directing recombination of a recombinant gene  
 CC back into the chromosome of the homologous species, replacing the native  
 CC copy of that gene; and (4) eliciting an immune response in an animal,  
 CC comprising separating an amino acid polymer comprising a recombinant Agfa  
 CC protein containing a replacement segment or segments of foreign amino  
 CC acid sequence or sequences grown on a Salmonella, E. coli or  
 CC Enterobacteriaceae host cell, from the host cell and introducing the  
 CC polymer into the animal in conjunction with a carrier or diluent. (I) is  
 CC useful for the expression of recombinant Agfa protein which is useful for  
 CC eliciting an immune response in an animal. In a fimbrial presentation  
 CC system the heterologous antigens are presented in high numbers (up to  
 CC 500,000 copies/cell), the hybrid fimbriin protein possesses both the  
 CC immunogenicity and adhesion properties relevant for an efficient live  
 CC vaccine, the carrier fimbrial subunit proteins are usually strong  
 CC immunogens, which may be important for directing an immune response  
 CC against the inserted epitope, and hybrid fimbriae are easy and  
 CC inexpensive to purify in large amount. The present sequence is given in  
 CC the exemplification of the present invention  
 XX  
 SQ Sequence 151 AA;  
 Query Match 90.3%; Score 700; DB 3; Length 151;  
 Best Local Similarity 89.9%; Pred. No. 4.2e-61;  
 Matches 143; Conservative 0; Mismatches 0; Indels 16; Gaps 2;  
 QY 1 MKLLKVAFAAIVVSGSALAGVVPQWGGGNGHNGGNSGPDSTLSIYQYGSANAALALQ 60  
 DB 1 MKLLKVAFAAIVVSGSALAGVVPQWGGGNGHNGGNSGPDSTLSIYQYGSANAALALQ 60  
 QY 61 SDARKSETTITQSGYNGADVGQADNSTIELTQNGFRNNATIDQWNAKNDQLVTRVVT 120  
 DB 61 SDARKSETTITQSGYNGADVGQADNSTIELTQNGFRNNATIDQWNAKNDQLVTRVVT 112  
 QY 121 HEMAHANOTASDSSVMVVRQVGFNNATANQY 151  
 DB 113 HEMAHAGGNAALVNQTSADSSVMVVRQVGFNNATANQY 151  
 RESULT 4  
 AAR74625  
 ID AAR74625 standard; protein; 151 AA.  
 XX  
 AC AAR74625;  
 XX  
 DT 25-MAR-2003 (revised)  
 DT 26-JUN-1995 (first entry)  
 XX  
 XX Agfa sequence.  
 XX Salmonella; Agfa; vaccine.  
 XX  
 OS Salmonella.

PN WO9425598-A2.  
 XX 10-NOV-1994.  
 XX 26-APR-1994; 94WO-IB000207.  
 XX 26-APR-1993; 93US-00054452.  
 XX (UYVI-) UNIV VICTORIA INNOVATION & DEV CORP.  
 PA (KING/) KING J.  
 XX Kay WW, Collinson SK, Clouthier SC, Doran JL;  
 XX WPI: 1994-358275/44.  
 DR N-PSDB; AAQ87467.  
 XX Eliciting an immune response to Salmonella - using attenuated Salmonella  
 PT strains, vector constructs, or compsns. contg. fimbrial type proteins.  
 XX Disclosure; Fig 7B; 95pp; English.  
 XX The Salmonella Agfa protein and DNA are used in vaccine and genetic  
 CC immunization compositions, respectively, to elicit an immune response to  
 CC Salmonella in animals (e.g. food producing animals) and humans. (Updated  
 CC on 25-MAR-2003 to correct PN field.)  
 XX Sequence 151 AA;  
 SQ Query Match 89.0%; Score 690; DB 2; Length 151;  
 Best Local Similarity 91.4%; Pred. No. 4.1e-60;  
 Matches 138; Conservative 2; Mismatches 11; Indels 0; Gaps 0;  
 QY 1 MKLLKVAFAAIVVSGSALAGVVPQWGGGNGHNGGNSGPDSTLSIYQYGSANAALALQ 60  
 DB 1 MKLLKVAFAAIVVSGSALAGVVPQWGGGNGHNGGNSGPDSTLSIYQYGSANAALALQ 60  
 QY 61 SDARKSETTITQSGYNGADVGQADNSTIELTQNGFRNNATIDQWNAKNDQLVTRVVT 120  
 DB 61 SDARKSETTITQSGYNGADVGQADNSTIELTQNGFRNNATIDQWNAKNDQLVTRVVT 120  
 QY 121 HEMAHANOTASDSSVMVVRQVGFNNATANQY 151  
 DB 121 NNAALVNQTSADSSVMVVRQVGFNNATANQY 151  
 RESULT 5  
 AAB36341  
 ID AAB36341 standard; protein; 151 AA.  
 XX  
 AC AAB36341;  
 XX  
 DT 26-FEB-2001 (first entry)  
 XX  
 DE Salmonella enteritidis Agfa amino acid sequence SEQ ID NO:5.  
 XX Salmonella; agfa; chromosomal gene replacement; fimbriin; epitope;  
 KW vaccine; immune response; immunogen.  
 XX Salmonella enteritidis.  
 OS  
 XX WO2000060102-A2.  
 PN  
 XX 12-OCT-2000.  
 XX  
 PF 05-APR-2000; 2000WO-CA000356.  
 XX  
 PR 05-APR-1999; 99US-0127888P.  
 XX  
 PA (UYVI-) UNIV VICTORIA.  
 XX White AP, Doran JL, Collison SK, Kay WW;  
 XX WPI: 2000-672631/65.

DR N-PSDB; AAC64617.  
 XX  
 PT Recombinant agfa gene having a segment replaced by a foreign DNA sequence  
 PT which encodes foreign epitope or antigen, expresses recombinant Agfa  
 PT protein useful for eliciting immune response in animal.  
 XX  
 PS Disclosure; Page 135; 139pp; English.  
 XX  
 CC The present invention describes a recombinant agfa gene (I) where a  
 CC segment of the gene has been replaced by a segment of a foreign DNA  
 CC sequence which encodes a foreign epitope or antigen. Also described are:  
 CC (1) use of thin aggregative fibrillae (SEF17/TAF) nucleation depended  
 CC assembly system of strains of Salmonella, Escherichia coli and  
 CC Enterobacteriaceae for the production of fibrillae comprising recombinant  
 CC Agfa, CsgA and Agfa-homologue fibrin subunits, respectively; (2)  
 CC directing recombination of a recombinant gene into the chromosome of the  
 CC homologous species; (3) directing recombination of a recombinant gene  
 CC back into the chromosome of the homologous species, replacing the native  
 CC copy of that gene; and (4) eliciting an immune response in an animal,  
 CC comprising separating an amino acid polymer comprising a recombinant Agfa  
 CC protein containing a replacement segment or segments of foreign amino  
 CC acid sequence or sequences grown on a Salmonella, E. coli or  
 CC Enterobacteriaceae host cell, from the host cell and introducing the  
 CC polymer into the animal in conjunction with a carrier or diluent. (I) is  
 CC useful for the expression of recombinant Agfa protein which is useful for  
 CC eliciting an immune response in an animal. In a fibrillar presentation  
 CC system the heterologous antigens are presented in high numbers (up to  
 CC 500,000 copies/cell), the hybrid fibrin protein possesses both the  
 CC immunogenicity and adhesion properties relevant for an efficient live  
 CC vaccine, the carrier fibrillar subunit proteins are usually strong  
 CC immunogens, which may be important for directing an immune response  
 CC against the inserted epitope, and hybrid fibrillae are easy and  
 CC inexpensive to purify in large amount. The present sequence is given in  
 CC the exemplification of the present invention  
 XX  
 SQ Sequence 151 AA;  
 Query Match 89.0%; Score 690; DB 3; Length 151;  
 Best Local Similarity 91.4%; Pred. No. 4.1e-60;  
 Matches 138; Conservative 2; Mismatches 11; Indels 0; Gaps 0;  
 QY 1 MKLLKVAAPAAIIVSGSALAGVVPQGGGNGGNSGGPDSITLSIYQYGSANAALALQ 60  
 DB 1 MKLLKVAAPAAIIVSGSALAGVVPQGGGNGGNSGGPDSITLSIYQYGSANAALALQ 60  
 QY 61 SDARKSETTITQSGYNGADVGQGDNSTIETQNGFRNNAIDQWNAKNDITVGVVVT 120  
 DB 61 SDARKSETTITQSGYNGADVGQGDNSTIETQNGFRNNAIDQWNAKNDITVGVVVT 120  
 QY 121 HEMAHANQASDSSVMVROVGFNNATANQY 151  
 DB 121 NNAALVNQASDSSVMVROVGFNNATANQY 151  
 RESULT 6  
 AAW23570  
 ID AAW23570 standard; protein; 151 AA.  
 XX  
 AC AAW23570;  
 DT 25-MAR-2003 (revised)  
 DT 29-SEP-1997 (first entry)  
 XX  
 DE Salmonella enteritidis 27655-3b agfa.  
 XX Enteropathogenic bacteria; enterobacteria; S. enteritidis; antibody.  
 XX Salmonella enteritidis.  
 OS  
 FH Key Location/Qualifiers  
 FT Misc-difference 123  
 FT /note= "Encoded by GCC"  
 XX

PN US5635617-A.  
 XX  
 PD 03-JUN-1997.  
 XX  
 PF 26-APR-1994; 94US-00233788.  
 XX  
 PR 26-APR-1993; 93US-00054452.  
 XX  
 PA (UYVI-) UNIV VICTORIA INNOVATION & DEV CORP.  
 XX  
 PI Collinson SK, Kay WW, Doran JL;  
 XX  
 XX WPI; 1997-309886/28.  
 DR N-PSDB; AAT74142.  
 XX  
 PT Isolated Salmonella gene agfa - used for diagnosis of Salmonella or  
 PT enteropathogenic bacteria of the Enterobacteria family.  
 XX  
 PS Example 2; Fig 7; 85pp; English.  
 XX  
 CC The present sequence represents agfa encoded by the full agfa gene  
 CC derived from Salmonella enteritidis 27655-3b. The nucleic acid can be  
 CC used to provide diagnostic assays for Salmonella and/or enteropathogenic  
 CC bacteria of the family Enterobacteria. It can also be used to provide  
 CC proteins and antibodies which can be used for assays. The nucleic acid  
 CC sequence can be used to provide probes or primers which can specifically  
 CC hybridise to nucleic acid molecules from greater than 99% of Salmonella  
 CC strains that are pathogenic to warm-blooded animals relative to nucleic  
 CC acid molecules from virtually all other microbial organisms. (Updated on  
 CC 25-MAR-2003 to correct PF field.)  
 XX  
 SQ Sequence 151 AA;  
 Query Match 88.9%; Score 689; DB 2; Length 151;  
 Best Local Similarity 91.4%; Pred. No. 5.2e-60;  
 Matches 138; Conservative 2; Mismatches 11; Indels 0; Gaps 0;  
 QY 1 MKLLKVAAPAAIIVSGSALAGVVPQGGGNGGNSGGPDSITLSIYQYGSANAALALQ 60  
 DB 1 MKLLKVAAPAAIIVSGSALAGVVPQGGGNGGNSGGPDSITLSIYQYGSANAALALQ 60  
 QY 61 SDARKSETTITQSGYNGADVGQGDNSTIETQNGFRNNAIDQWNAKNDITVGVVVT 120  
 DB 61 SDARKSETTITQSGYNGADVGQGDNSTIETQNGFRNNAIDQWNAKNDITVGVVVT 120  
 QY 121 HEMAHANQASDSSVMVROVGFNNATANQY 151  
 DB 121 NNPALVNQASDSSVMVROVGFNNATANQY 151  
 RESULT 7  
 AAB36346  
 ID AAB36346 standard; protein; 151 AA.  
 XX  
 AC AAB36346;  
 DT 26-FEB-2001 (first entry)  
 XX  
 DE Agfa::PT3#1 amino acid sequence SEQ ID NO:12.  
 XX  
 KW Salmonella; agfa; chromosomal gene replacement; fimbria; epitope;  
 KW vaccine; immune response; immunogen.  
 XX  
 OS Salmonella enteritidis.  
 OS Escherichia coli.  
 OS Synthetic.  
 XX  
 PN WO2000060102-A2.  
 XX  
 PD 12-OCT-2000.  
 XX  
 PF 05-APR-2000; 2000WO-CA000356.  
 XX

```

PR 05-APR-1999; 9905-0127888P.
PA (UYVI-) UNIV VICTORIA.
XX
PI White AP, Doran JL, Collison SK, Kay WW;
XX
DR WPI: 2000-672631/65.
XX N-PSDB; AAC64622.
XX
PT Recombinant agfa gene having a segment replaced by a foreign DNA sequence
PT which encodes foreign epitope or antigen, expresses recombinant Agfa
PT protein useful for eliciting immune response in animal.
XX
PS Disclosure; Page 135; 139pp; English.
XX
XX The present invention describes a recombinant agfa gene (I) where a
CC segment of the gene has been replaced by a segment of a foreign DNA
CC sequence which encodes a foreign epitope or antigen. Also described are:
CC (1) use of thin aggregative fimbriae (SEF17/RAF) nucleation depended
CC assembly system of strains of Salmonella, Escherichia coli and
CC Enterobacteriaceae for the production of fimbriae comprising recombinant
CC Agfa, CsgA and Agfa-homologue fimbria subunits, respectively; (2)
CC directing recombination of a recombinant gene into the chromosome of the
CC homologous species; (3) directing recombination of a recombinant gene
CC back into the chromosome of the homologous species, replacing the native
CC copy of that gene; and (4) eliciting an immune response in an animal,
CC comprising separating an amino acid polymer comprising a recombinant Agfa
CC protein containing a replacement segment or segments of foreign amino
CC acid sequence or sequences grown on a Salmonella, E. coli or
CC Enterobacteriaceae host cell, from the host cell and introducing the
CC polymer into the animal in conjunction with a carrier or diluent. (I) is
CC useful for the expression of recombinant Agfa protein which is useful for
CC eliciting an immune response in an animal. In a fimbrial presentation
CC system the heterologous antigens are presented in high numbers (up to
CC 500,000 copies/cell), the hybrid fimbria protein possesses both the
CC immunogenicity and adhesion properties relevant for an efficient live
CC vaccine, the carrier fimbrial subunit proteins are usually strong
CC immunogens, which may be important for directing an immune response
CC against the inserted epitope, and hybrid fimbriae are easy and
CC inexpensive to purify in large amount. The present sequence is given in
CC the exemplification of the present invention
XX
XX Sequence 151 AA;
SQ
Query Match 84.5%; Score 655; DB 3; Length 151;
Best Local Similarity 79.8%; Pred. No. 1.2e-56;
Matches 134; Conservative 0; Mismatches 0; Indels 34; Gaps 2
QY 1 MKLLKVAFAAIYVSGSALAGVVPQGGGHHNGGNSGDPSTLSIYQVGSANAALAQ 60
DB 1 MKLLKVAFAAIYVSGSALAGVVPQGGGHHNGGNSGDPSTLSIYQVGSANAALAQ 60
QY 61 SDARKSETTITQSGYNGADVGCGADNSTIELTQGFRRNATIDQWNAK----- 109
DB 61 SDARKSETTITQSGYNGADVGCGADNSTIELTQGFRRNATIDQWNAKNSDITVGYGG 120
QY 110 -----NYDQLVTRVVTHEMAHANQTASDSSVWVQVQFGNNATANQY 151
DB 121 NNAALVNYDQLVTRVVTHEMAHA-----NNATANQY 151
RESULT 8
AAB36353
ID AAB36353 standard; protein; 151 AA.
XX
XX AAB36353;
XX
XX 26-FEB-2001 (first entry)
XX
XX Agfa::PT3#48 amino acid sequence SEQ ID NO:26.
XX
XX Salmonella; agfa; chromosomal gene replacement; fimbria; epitope;
XX vaccine; immune response; immunogen.
KW

```

```

XX OS Salmonella enteritidis.
OS Escherichia coli.
OS Synthetic.
XX PN WO200060102-A2.
XX PD 12-OCT-2000.
XX PF 05-APR-2000; 2000WO-CA000356.
XX PR 05-APR-1999; 99US-0127888P.
XX PA (UWI-) UNIV VICTORIA.
XX PI White AP, Doran JL, Collison SK, Kay WW;
XX WPI; 2000-672631/65.
XX DR N-PSDB; AAC64629.
XX PT Recombinant agfa gene having a segment replaced by a foreign DNA sequence
PT which encodes foreign epitope or antigen, expresses recombinant Agfa
PT protein useful for eliciting immune response in animal.
XX Disclosure; Page 138; 139pp; English.
XX The present invention describes a recombinant agfa gene (I) where a
CC segment of the gene has been replaced by a segment of a foreign DNA
CC sequence which encodes a foreign epitope or antigen. Also described are:
CC (1) use of thin aggregative fimbriae (SEF17/TAF) nucleation depended
CC assembly system of strains of Salmonella, Escherichia coli and
CC Enterobacteriaceae for the production of fimbriae comprising recombinant
CC Agfa, CsgA and Agfa-homologue fimbryn subunits, respectively; (2)
CC directing recombination of a recombinant gene into the chromosome of the
CC homologous species; (3) directing recombination of a recombinant gene
CC back into the chromosome of the homologous species, replacing the native
CC copy of that gene; and (4) eliciting an immune response in an animal,
CC comprising separating an amino acid polymer comprising a recombinant Agfa
CC protein containing a replacement segment or segments of foreign amino
CC acid sequence or sequences grown on a Salmonella, E. coli or
CC Enterobacteriaceae host cell, from the host cell and introducing the
CC polymer into the animal in conjunction with a carrier or diluent. (I) is
CC useful for the expression of recombinant Agfa protein which is useful for
CC eliciting an immune response in an animal. In a fimbrial presentation
CC system the heterologous antigens are presented in high numbers (up to
CC 500,000 copies/cell), the hybrid fimbryn protein possesses both the
CC immunogenicity and adhesion properties relevant for an efficient live
CC vaccine, the carrier fimbrial subunit proteins are usually strong
CC immunogens, which may be important for directing an immune response
CC against the inserted epitope, and hybrid fimbriae are easy and
CC inexpensive to purify in large amount. The present sequence is given in
CC the exemplification of the present invention
XX Sequence 151 AA;
SQ Query Match 78.6%; Score 609; DB 3; Length 151;
Best Local Similarity 82.1%; Pred. No. 4.2e-52;
Matches 124; Conservative 4; Mismatches 23; Indels 0; Gaps 0
QY 1 MKLLKVAFAAIWSSGALAGVVPQWGCGNGNHGGNSGPDSTLSIYOYGSAALALQ 60
DB 1 MKLLKVAFAAIWSSGALAGVVPQWGCGNGNHGGNSGPDSTLSIYOYGSAALAYDQ 60
QY 61 SPARKSETTITSGYGNADVCQGADNSTIELTQTONGFRNNATIDQNAKNVDQLVTRVVT 120
DB 61 LVTRVVTVEHMAHYGNADVCQGADNSTIELTQTONGFRNNATIDQNAKNSDITVGQYGG 120
QY 121 HEMAHANOTASDSSVMVRQVGFENNATANQY 151
DB 121 NNAALVNQTASDSSVMVRQVGFENNATANQY 151

```

AAB36351  
 ID AAB36351 standard; protein; 151 AA.  
 AC AAB36351;  
 XX  
 DT 26-FEB-2001 (first entry)  
 XX  
 DE Agfa::PT3#6 amino acid sequence SEQ ID NO:22.  
 XX  
 KW Salmonella; agfa; chromosomal gene replacement; fimbrin; epitope;  
 XX vaccine; immune response; immunogen.  
 OS Salmonella enteritidis.  
 OS Escherichia coli.  
 OS Synthetic.  
 XX  
 PN WO200060102-A2.  
 XX  
 XX 12-OCT-2000.  
 XX  
 PF 05-APR-2000; 2000WO-CA000356.  
 XX  
 XX 05-APR-1999; 99US-0127888P.  
 XX  
 PA (UYVI-) UNIV VICTORIA.  
 XX  
 PI White AP, Doran JL, Collison SK, Kay WW;  
 XX  
 DR WPI; 2000-672631/65.  
 DR N-PSDB; AAC64625.  
 XX  
 XX Recombinant agfa gene having a segment replaced by a foreign DNA sequence  
 PT which encodes foreign epitope or antigen, expresses recombinant Agfa  
 PT protein useful for eliciting immune response in animal.  
 XX  
 PS Disclosure; Page 137; 139pp; English.  
 XX  
 CC The present invention describes a recombinant agfa gene (I) where a  
 CC segment of the gene has been replaced by a segment of a foreign DNA  
 CC sequence which encodes a foreign epitope or antigen. Also described are:  
 CC (1) use of thin aggregative fimbriae (SEF17/TAF) nucleation depended  
 CC assembly system of strains of Salmonella, Escherichia coli and  
 CC Enterobacteriaceae for the production of fimbriae comprising recombinant  
 CC Agfa, CsgA and Agfa-homologue fimbrial subunits, respectively; (2)  
 CC directing recombination of a recombinant gene into the chromosome of the  
 CC homologous species; (3) directing recombination of a recombinant gene  
 CC back into the chromosome of the homologous species, replacing the native  
 CC copy of that gene; and (4) eliciting an immune response in an animal,  
 CC comprising separating an amino acid polymer comprising a recombinant Agfa  
 CC protein containing a replacement segment or segments of foreign amino  
 CC acid sequence or sequences grown on a Salmonella, E. coli or  
 CC Enterobacteriaceae host cell, from the host cell and introducing the  
 CC polymer into the animal in conjunction with a carrier or diluent. (I) is  
 CC useful for the expression of recombinant Agfa protein which is useful for  
 CC eliciting an immune response in an animal. In a fimbrial presentation  
 CC system the heterologous antigens are presented in high numbers (up to  
 CC 500,000 copies/cell), the hybrid fimbrial protein possesses both the  
 CC immunogenicity and adhesion properties relevant for an efficient live  
 CC vaccine, the carrier fimbrial subunit proteins are usually strong  
 CC immunogens, which may be important for directing an immune response  
 CC against the inserted epitope, and hybrid fimbriae are easy and  
 CC inexpensive to purify in large amount. The present sequence is given in  
 CC the exemplification of the present invention  
 XX  
 XX Sequence 151 AA;  
 Query Match 78.6%; Score 609; DB 3; Length 151;  
 Best Local Similarity 73.6%; Pred. No. 4.2e-52;  
 Matches 128; Conservative 0; Mismatches 0; Indels 46; Gaps 2;  
 1 MKLLKVAFAAIVWGSALAGVVPQWGGNGNGGNSGPDSTLSTIYQGSANAALQ 60  
 1 MKLLKVAFAAIVWGSALAGVVPQWGGNGNGGNSGPDSTLSTIYQGSANAALQ 60

QY 61 SDARKSETTITQSGYNGADVGQAGDNSTIELTQGFNNATIDQMNKYNQDLVTRVVT 120  
 DB ||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
 61 SDARKSETTITQSGYNGADVGQAGDNSTIELTQGFNNATIDQMNKYNQDLVTRVVT 97  
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
 QY 121 HEMAHA-----NOTASDSSVMYRVQVGFNNATANQY 151  
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
 DB 98 HEMAHADQMNKYNQDLVTRVVT-----NOTASDSSVMYRVQVGFNNATANQY 151  
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||

RESULT 10  
 AAB36349  
 ID AAB36349 standard; protein; 151 AA.  
 XX  
 AC AAB36349;  
 XX  
 DT 26-FEB-2001 (first entry)  
 XX  
 DE Agfa::PT3#4 amino acid sequence SEQ ID NO:18.  
 XX  
 KW Salmonella; agfa; chromosomal gene replacement; fimbrin; epitope;  
 KW vaccine; immune response; immunogen.  
 XX  
 XX Salmonella enteritidis.  
 OS Escherichia coli.  
 OS Synthetic.  
 XX  
 PN WO200060102-A2.  
 XX  
 XX 12-OCT-2000.  
 XX  
 PF 05-APR-2000; 2000WO-CA000356.  
 XX  
 XX 05-APR-1999; 99US-0127888P.  
 PR  
 XX (UYVI-) UNIV VICTORIA.  
 PA  
 PI White AP, Doran JL, Collison SK, Kay WW;  
 XX  
 DR WPI; 2000-672631/65.  
 DR N-PSDB; AAC64625.  
 XX  
 PT Recombinant agfa gene having a segment replaced by a foreign DNA sequence  
 PT which encodes foreign epitope or antigen, expresses recombinant Agfa  
 PT protein useful for eliciting immune response in animal.  
 XX  
 PS Disclosure; Page 136; 139pp; English.

CC The present invention describes a recombinant agfa gene (I) where a  
 CC segment of the gene has been replaced by a segment of a foreign DNA  
 CC sequence which encodes a foreign epitope or antigen. Also described are:  
 CC (1) use of thin aggregative fimbriae (SEF17/TAF) nucleation depended  
 CC assembly system of strains of Salmonella, Escherichia coli and  
 CC Enterobacteriaceae for the production of fimbriae comprising recombinant  
 CC Agfa, CsgA and Agfa-homologue fimbrial subunits, respectively; (2)  
 CC directing recombination of a recombinant gene into the chromosome of the  
 CC homologous species; (3) directing recombination of a recombinant gene  
 CC back into the chromosome of the homologous species, replacing the native  
 CC copy of that gene; and (4) eliciting an immune response in an animal,  
 CC comprising separating an amino acid polymer comprising a recombinant Agfa  
 CC protein containing a replacement segment or segments of foreign amino  
 CC acid sequence or sequences grown on a Salmonella, E. coli or  
 CC Enterobacteriaceae host cell, from the host cell and introducing the  
 CC polymer into the animal in conjunction with a carrier or diluent. (I) is  
 CC useful for the expression of recombinant Agfa protein which is useful for  
 CC eliciting an immune response in an animal. In a fimbrial presentation  
 CC system the heterologous antigens are presented in high numbers (up to  
 CC 500,000 copies/cell), the hybrid fimbrial protein possesses both the  
 CC immunogenicity and adhesion properties relevant for an efficient live  
 CC vaccine, the carrier fimbrial subunit proteins are usually strong  
 CC immunogens, which may be important for directing an immune response  
 CC against the inserted epitope, and hybrid fimbriae are easy and  
 CC inexpensive to purify in large amount. The present sequence is given in



The present invention describes a recombinant agfa gene (I) where a segment of the gene has been replaced by a segment of a foreign DNA sequence which encodes a foreign epitope or antigen. Also described are: (1) use of thin aggregative fibrillae (SFA17/TAF) nucleation depended assembly system of strains of *Salmonella*, *Escherichia coli* and *Enterobacteriaceae* for the production of fimbriae comprising recombinant Agfa, CsgA and Agfa-homologue fibrin subunits, respectively; (2) directing recombination of a recombinant gene into the chromosome of the homologous species; (3) directing recombination of a recombinant gene back into the chromosome of the homologous species, replacing the native copy of that gene; and (4) eliciting an immune response in an animal, comprising separating an amino acid polymer comprising a recombinant Agfa protein containing a replacement segment or segments of foreign amino acid sequence or sequences grown on a *Salmonella*, *E. coli* or

CC (1) use of thin aggregative fimbriae (SEF17/TAF) nucleation depended  
 CC assembly system of strains of Salmonella, Escherichia coli and  
 CC Enterobacteriaceae for the production of fimbriae comprising recombinant  
 CC AgfA, CsgA and AgfA-homologue fimbrial subunits, respectively; (2)  
 CC directing recombination of a recombinant gene into the chromosome of the  
 CC homologous species; (3) directing recombination of a recombinant gene  
 CC back into the chromosome of the homologous species, replacing the native  
 CC copy of that gene; and (4) eliciting an immune response in an animal,  
 CC comprising separating an amino acid polymer comprising a recombinant AgfA  
 CC protein containing a replacement segment or segments of foreign amino  
 CC acid sequence or sequences grown on a Salmonella, E. coli or  
 CC Enterobacteriaceae host cell, from the host cell and introducing the  
 CC polymer into the animal in conjunction with a carrier or diluent. (1) is  
 CC useful for the expression of recombinant AgfA protein which is useful for  
 CC eliciting an immune response in an animal. In a fimbrial presentation  
 CC system the heterologous antigens are presented in high numbers (up to  
 CC 500,000 copies/cell), the hybrid fimbrial protein possesses both the  
 CC immunogenicity and adhesion properties relevant for an efficient live  
 CC vaccine, the carrier fimbrial subunit proteins are usually strong  
 CC immunogens, which may be important for directing an immune response  
 CC against the inserted epitope, and hybrid fimbriae are easy and  
 CC inexpensive to purify in large amount. The present sequence is given in  
 CC the exemplification of the present invention

XX Sequence 151 AA;

Query Match 77.4%; Score 600; DB 3; Length 151;  
 Best Local Similarity 81.5%; Pred. No. 3.2e-51;  
 Matches 123; Conservative 3; Mismatches 25; Indels 0; Gaps 0;  
 QY 1 MKLLKVAAPAAIIVSGSALAGVVPQWGGGNGGNGSGSPDSTLSIYQGSANAALALQ 60  
 DB 1 MKLLKVAAPAAIIVSGSALAGVVPQWGGGNGGNGSGSPDSTLSIYQGSANAALALQ 60  
 QY 61 SDARKSETTITQSGYNGADYVQGGADNSTIELTQNGFRNNATIDQWNAKNDQLVTRVVT 120  
 DB 61 SDARKSETTITQSGYNGADYVQGGADNSTIELTQNGFRNNATIDQWNAKNDQLVTRVVT 120

QY 121 HEMAHANQTASDSSVMVRQVGFNNATANQY 151  
 DB 121 NNAALVNQTASDSSVMVRQVGFNNATANQY 151

RESULT 13  
 AAB36348  
 ID AAB36348 standard; protein; 151 AA.  
 AC AAB36348;

DT 26-FEB-2001 (first entry)

XX AgfA::PT3#3 amino acid sequence SEQ ID NO:16.

DE Salmonella; agfA; chromosomal gene replacement; fimbrin; epitope;  
 KW vaccine; immune response; immunogen.

XX Salmonella enteritidis.

OS Escherichia coli.

OS Synthetic.

XX WO200060102-A2.

PN 12-OCT-2000.

XX 05-APR-2000; 2000WO-CA000356.

PF 05-APR-1999; 99US-0127888P.

XX (UYVI-) UNIV VICTORIA.

XX White AP, Doran JL, Collison SK, Kay WW;

XX WPI; 2000-672631/65.

DR N-PSDB; AAC64624.  
 XX Recombinant agfA gene having a segment replaced by a foreign DNA sequence  
 PT which encodes foreign epitope or antigen, expresses recombinant AgfA  
 PT protein useful for eliciting immune response in animal.  
 XX Disclosure; Page 136; 139pp; English.

CC The present invention describes a recombinant agfA gene (1) where a  
 CC segment of the gene has been replaced by a segment of a foreign DNA  
 CC sequence which encodes a foreign epitope or antigen. Also described are:  
 CC (1) use of thin aggregative fimbriae (SEF17/TAF) nucleation depended  
 CC assembly system of strains of Salmonella, Escherichia coli and  
 CC Enterobacteriaceae for the production of fimbriae comprising recombinant  
 CC AgfA, CsgA and AgfA-homologue fimbrial subunits, respectively; (2)  
 CC directing recombination of a recombinant gene into the chromosome of the  
 CC homologous species; (3) directing recombination of a recombinant gene  
 CC back into the chromosome of the homologous species, replacing the native  
 CC copy of that gene; and (4) eliciting an immune response in an animal, AgfA  
 CC protein containing a replacement segment or segments of foreign amino  
 CC acid sequence or sequences grown on a Salmonella, E. coli or  
 CC Enterobacteriaceae host cell, from the host cell and introducing the  
 CC polymer into the animal in conjunction with a carrier or diluent. (1) is  
 CC useful for the expression of recombinant AgfA protein which is useful for  
 CC eliciting an immune response in an animal. In a fimbrial presentation  
 CC system the heterologous antigens are presented in high numbers (up to  
 CC 500,000 copies/cell), the hybrid fimbrial protein possesses both the  
 CC immunogenicity and adhesion properties relevant for an efficient live  
 CC vaccine, the carrier fimbrial subunit proteins are usually strong  
 CC immunogens, which may be important for directing an immune response  
 CC against the inserted epitope, and hybrid fimbriae are easy and  
 CC inexpensive to purify in large amount. The present sequence is given in  
 CC the exemplification of the present invention

XX Sequence 151 AA;

Query Match 74.2%; Score 575; DB 3; Length 151;  
 Best Local Similarity 81.5%; Pred. No. 9.5e-49;  
 Matches 123; Conservative 5; Mismatches 23; Indels 0; Gaps 0;

QY 1 MKLLKVAAPAAIIVSGSALAGVVPQWGGGNGGNGSGSPDSTLSIYQGSANAALALQ 60  
 DB 1 MKLLKVAAPAAIIVSGSALAGVVPQWGGGNGGNGSGSPDSTLSIYQGSANAALALQ 60

QY 61 SDARKSETTITQSGYNGADYVQGGADNSTIELTQNGFRNNATIDQWNAKNDQLVTRVVT 120  
 DB 61 SDARKSETTITQSGYNGADYVQGGADNSTIELTQNGFRNNATIDQWNAKNDQLVTRVVT 120

QY 121 HEMAHANQTASDSSVMVRQVGFNNATANQY 151

DB 121 NNAALVNQTASDSSVMVRQVGFNNATANQY 151

RESULT 14  
 AAB36343

ID AAB36343 standard; protein; 151 AA.

AC AAB36343;

DT 26-FEB-2001 (first entry)

XX Escherichia coli CsgA amino acid sequence SEQ ID NO:7.

DE Salmonella; agfA; chromosomal gene replacement; fimbrin; epitope;  
 KW vaccine; immune response; immunogen.

XX Escherichia coli.

OS WO200060102-A2.

XX 12-OCT-2000.

PF 05-APR-2000; 2000WO-CA000356.  
 XX 05-APR-1999; 99US-0127888P.  
 XX (UYVI-) UNIV VICTORIA.  
 XX White AP, Doran JL, Collison SK, Kay WW;  
 XX WPI; 2000-672631/65.  
 DR N-PSDB; AAC64619.  
 XX Recombinant agfA gene having a segment replaced by a foreign DNA sequence  
 PT which encodes foreign epitope or antigen, expresses recombinant AgfA  
 PT protein useful for eliciting immune response in animal.  
 XX Disclosure; Page 135; 139pp; English.  
 XX The present invention describes a recombinant agfA gene (I) where a  
 CC segment of the gene has been replaced by a segment of a foreign DNA  
 CC sequence which encodes a foreign epitope or antigen. Also described are:  
 CC (1) use of thin aggregative fimbriae (SEF17/TAF) nucleation depended  
 CC assembly system of strains of Salmonella, Escherichia coli and  
 CC Enterobacteriaceae for the production of fimbriae comprising recombinant  
 CC AgfA, CsgA and AgfA-homologue fimbrial subunits, respectively; (2)  
 CC directing recombination of a recombinant gene into the chromosome of the  
 CC homologous species; (3) directing recombination of a recombinant gene  
 CC back into the chromosome of the homologous species, replacing the native  
 CC copy of that gene; and (4) eliciting an immune response in an animal,  
 CC comprising separating an amino acid polymer comprising a recombinant AgfA  
 CC protein containing a replacement segment or segments of foreign amino  
 CC acid sequence or sequences grown on a Salmonella, E. coli or  
 CC Enterobacteriaceae host cell, from the host cell and introducing the  
 CC polymer into the animal in conjunction with a carrier or diluent. (I) is  
 CC useful for the expression of recombinant AgfA protein which is useful for  
 CC eliciting an immune response in an animal. In a fimbrial presentation  
 CC system the heterologous antigens are presented in high numbers (up to  
 CC 500,000 copies/cell), the hybrid fimbrial protein possesses both the  
 CC immunogenicity and adhesion properties relevant for an efficient live  
 CC vaccine, the carrier fimbrial subunit proteins are usually strong  
 CC immunogens, which may be important for directing an immune response  
 CC against the inserted epitope, and hybrid fimbriae are easy and  
 CC inexpensive to purify in large amount. The present sequence is given in  
 CC the exemplification of the present invention  
 XX Sequence 151 AA;  
 SQ Query Match 68.8%; Score 533; DB 3; Length 151;  
 Best Local Similarity 70.2%; Pred. No. 1.3e-44;  
 Matches 106; Conservative 18; Mismatches 27; Indels 0; Gaps 0;  
 QY 1 MKLLKVAAPAAIVVSGSALAGVVPQWGGGNGGNSGDPSTLSIYQYGSANAALALQ 60  
 DB 1 MKLLKVAATAAIVFSGSALAGVVPQYGGGNGGNSGPNSELNIYQYGGNSALALQ 60  
 QY 61 SDARKSETTITQSGYNGADVGADNSTIELTQNGFRNNATIDOWNAKNYDQLVTRVVT 120  
 DB 61 TDARNSDLTITQGGNGADVGQSDSSIDLTLQRFNGSATLDQWNGKNSMTVKQFGG 120  
 QY 121 HEMAHANQTASDSSVMVRYQVFGNNATANQY 151  
 DB 121 GNGAAVDQTASNSNVNTQVFGNNATAHQY 151  
 Search completed: August 2, 2004, 14:48:27  
 Job time: 44.9 secs

KW Plasma protein; immune response; antibacterial; vaccine; gene therapy.  
 XX Escherichia coli.  
 OS WO2003064446-A2.  
 XX 07-AUG-2003.  
 PD 30-JAN-2003; 2003WO-EP000943.  
 XX 31-JAN-2002; 2002GB-00002275.  
 XX (HANS-) HANSA MEDICAL RES AB.  
 XX Bjoerck L, Olsen A, Wikstroem M, Herwald H;  
 PI WPI; 2003-646136/61.  
 XX N-PSDB; ACF36153.  
 DR New isolated peptide capable of binding a mammalian plasma protein,  
 XX useful in the manufacture of a medicament for the prevention and/or  
 PT treatment of a bacterial infection, such as Escherichia coli, Salmonella  
 PT or Shigella infections.  
 XX Disclosure; Page 41-42; 42pp; English.  
 PS The invention relates to an isolated peptide capable of binding a  
 CC mammalian plasma protein or of generating an immune response in a mammal  
 CC selected from sequences shown in ABR82642, ABR82648-49. The peptide or  
 CC antibody is useful for treating a bacterial infection in a human or  
 CC animal or in the manufacture of a medicament for the prophylactic  
 CC treatment of a bacterial infection, such as Escherichia coli, Salmonella  
 CC or Shigella infection. The peptide that is immobilized on a solid support  
 CC is also useful as a reagent for determining the ability of a plasma  
 CC protein to bind to bacteria. The present sequence represents an E. coli  
 CC 15 kDa protein  
 XX Sequence 151 AA;  
 SQ Query Match 68.1%; Score 528; DB 7; Length 151;  
 Best Local Similarity 69.5%; Pred. No. 4.2e-44;  
 Matches 105; Conservative 18; Mismatches 28; Indels 0; Gaps 0;  
 QY 1 MKLLKVAAPAAIVVSGSALAGVVPQWGGGNGGNSGDPSTLSIYQYGSANAALALQ 60  
 DB 1 MKLLKVAATAAIVFSGSALAGVVPQYGGGNGGNSGPNSELNIYQYGGNSALALQ 60  
 QY 61 SDARKSETTITQSGYNGADVGQSDNSTIELTQNGFRNNATIDOWNAKNYDQLVTRVVT 120  
 DB 61 TDARNSDLTITQGGNGADVGQSDSSIDLTLQRFNGSATLDQWNGKNSMTVKQFGG 120  
 QY 121 HEMAHANQTASDSSVMVRYQVFGNNATANQY 151  
 DB 121 GNGAAVDQTASNSNVNTQVFGNNATAHQY 151  
 Search completed: August 2, 2004, 14:48:27  
 Job time: 44.9 secs

This page Blank (uspto)



Db 1 MLLKVAFAAIVVSGSALAGVVPQWGGGNGHNGSGPDSSTLSIYQGSANAALALQ 60  
 QY 61 SPARKSETTITGSGYNGADVGQGDNSTIELTQNGFRNNATIDQWNAKNDQLVTRVVT 120  
 Db 61 SPARKSETTITGSGYNGADVGQGDNSTIELTQNGFRNNATIDQWNAKNDQLVTRVVT 120  
 QY 121 HEMAHANQASDSSVMVRQVGFNNATANQY 151  
 Db 121 NNPALVNTQASDSSVMVRQVGFNNATANQY 151

RESULT 2  
 US-08-233-788A-57  
 ; Sequence 57, Application US/08233788A  
 ; Patent No. 5635617  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Doran, James L.  
 ; APPLICANT: Kay, William W.  
 ; APPLICANT: Collinson, Karen S.  
 ; APPLICANT: Clouthier, Sharon C.  
 ; TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR DETECTION  
 ; TITLE OF INVENTION: OF SALMONELLA  
 ; NUMBER OF SEQUENCES: 61  
 ; CORRESPONDENCE ADDRESS:  
 ; ADDRESSEE: Seed and Berry  
 ; STREET: 6300 Columbia Center, 701 Fifth Avenue  
 ; CITY: Seattle  
 ; STATE: Washington  
 ; COUNTRY: U.S.A.  
 ; ZIP: 98104-7092  
 ; COMPUTER READABLE FORM:  
 ; MEDIUM TYPE: Floppy disk  
 ; COMPUTER: IBM PC compatible  
 ; OPERATING SYSTEM: PC-DOS/MS-DOS  
 ; SOFTWARE: Patent In Release #1.0, Version #1.25  
 ; CURRENT APPLICATION DATA:  
 ; APPLICATION NUMBER: US/08/233,788A  
 ; FILING DATE: 26-APR-1994  
 ; CLASSIFICATION: 435  
 ; ATTORNEY/AGENT INFORMATION:  
 ; NAME: King, Joshua  
 ; REGISTRATION NUMBER: 35,570  
 ; REFERENCE/DOCKET NUMBER: 920043.403C2  
 ; TELEPHONE: (206) 622-4900  
 ; TELEFAX: (206) 682-6031  
 ; INFORMATION FOR SEQ ID NO: 57:  
 ; SEQUENCE CHARACTERISTICS:  
 ; LENGTH: 120 amino acids  
 ; TYPE: amino acid  
 ; TOPOLOGY: linear  
 ; MOLECULE TYPE: protein  
 ; US-08-233-788A-57

Query Match 65.2%; Score 505; DB 1; Length 120;  
 Best Local Similarity 88.4%; Pred. No. 1.7e-42;  
 Matches 99; Conservative 2; Mismatches 11; Indels 0; Gaps 0;  
 QY 22 VVPQWGGGNGHNGSGPDSSTLSIYQGSANAALQSDARKSETTITGSGYNGADV 81  
 Db 1 VVPQWGGGNGHNGSGPDSSTLSIYQGSANAALQSDARKSETTITGSGYNGADV 60  
 QY 82 GQGDNSTIELTQNGFRNNATIDQWNAKNDQLVTRVVTHEMAHANQASDS 133  
 Db 61 GQGDNSTIELTQNGFRNNATIDQWNAKNDQLVTRVVTGQGNNAALVNTQASDS 112

RESULT 3  
 US-09-196-387-8  
 ; Sequence 8, Application US/09196387  
 ; Patent No. 6277613  
 ; GENERAL INFORMATION:

; APPLICANT: de Lange, Titia  
 ; APPLICANT: Smith, Susan  
 ; TITLE OF INVENTION: A PROTEIN THAT BINDS TO TRF1 AND METHODS  
 ; TITLE OF INVENTION: OF USE THEREOF  
 ; NUMBER OF SEQUENCES: 12  
 ; CORRESPONDENCE ADDRESS:  
 ; ADDRESSEE: Klauber & Jackson  
 ; STREET: 411 Hackensack Avenue, 4th Floor  
 ; CITY: Hackensack  
 ; STATE: New Jersey  
 ; COUNTRY: USA  
 ; ZIP: 07601

; COMPUTER READABLE FORM:  
 ; MEDIUM TYPE: Floppy disk  
 ; COMPUTER: IBM PC compatible  
 ; OPERATING SYSTEM: PC-DOS/MS-DOS  
 ; SOFTWARE: Patent In Release #1.0, Version #1.30  
 ; CURRENT APPLICATION DATA: US/09/196,387  
 ; FILING DATE:  
 ; CLASSIFICATION:  
 ; PRIOR APPLICATION DATA:  
 ; APPLICATION NUMBER: 09/095,225  
 ; FILING DATE: June 10, 1998  
 ; ATTORNEY/AGENT INFORMATION:  
 ; NAME: Jackson Esq., David A.  
 ; REGISTRATION NUMBER: 26,742  
 ; REFERENCE/DOCKET NUMBER: 600-1-230 CIP1  
 ; TELECOMMUNICATION INFORMATION:  
 ; TELEPHONE: 201-487-5800  
 ; TELEFAX: 201-343-1684  
 ; TELEX: 133521

; INFORMATION FOR SEQ ID NO: 8:  
 ; SEQUENCE CHARACTERISTICS:  
 ; LENGTH: 673 amino acids  
 ; TYPE: amino acid  
 ; TOPOLOGY: linear  
 ; MOLECULE TYPE: protein  
 ; US-09-196-387-8

Query Match 11.7%; Score 90.5; DB 3; Length 673;  
 Best Local Similarity 28.4%; Pred. No. 1.1;  
 Matches 42; Conservative 16; Mismatches 61; Indels 29; Gaps 6;

QY 6 VAAFAAI-VVGSALAGVVPQWGGGNGHNGSGPDSSTLSIYQGSANAALQSDAR 64  
 Db 99 VAAAPVPAVSTSSAAGVAPNPAGSGNNSPSSSFTSS-SSSSPSPGSSLAESPEAA 157  
 QY 65 KSETTIT----QSGYNGADVGQGDNSTIELTQNGFRNNATIDQWNAKNDQLVTRVVT 120  
 Db 158 GVSSTAPLGPAGPGTGVPAVSGALRELLEACRNG-----DVSRLKRLVDA--- 204  
 QY 121 HEMAHANQASD-----SSVMVRQVGF 143  
 Db 205 -----ANVNAKDMAGRKSPPLHFAAGF 227

RESULT 4  
 US-09-841-835-8  
 ; Sequence 8, Application US/09841835  
 ; Patent No. 650587  
 ; GENERAL INFORMATION:  
 ; APPLICANT: de Lange, Titia  
 ; APPLICANT: Smith, Susan

; TITLE OF INVENTION: A PROTEIN THAT BINDS TO TRF1 AND METHODS  
 ; TITLE OF INVENTION: OF USE THEREOF  
 ; NUMBER OF SEQUENCES: 12  
 ; CORRESPONDENCE ADDRESS:  
 ; ADDRESSEE: Klauber & Jackson  
 ; STREET: 411 Hackensack Avenue, 4th Floor  
 ; CITY: Hackensack  
 ; STATE: New Jersey  
 ; COUNTRY: USA



```

; SEQUENCE CHARACTERISTICS:
; LENGTH: 949 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-09-841-835-10

```

```

Query Match      11.7%; Score 90.5; DB 4; Length 949;
Best Local Similarity 28.4%; Pred. No. 1.8;
Matches         42; Conservative 16; Mismatches 61; Indels 29; Gaps 6;

QY    6 VAAFAAI-VVGSALAGVVPQWGGNGNHGNSGPDSTLSIYVGSAHALAQSDAR 64
       ||||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| :
Db     99 VAAAPVVPVAVSTSSAAGVAPNPAGSGNNPSFSSSPTSS-SSSPSPGSSLAESPEAA 157
       ||||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| :
QY    65 KSETIT-----QSGYGNGADYGCQGDANSTTELQNGFRNATIDOWNAKNYDQLVTRVVT 120
       ||||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| :
Db     158 GYSSTAPLPGGAAGPCTGVPAVSGALZELLEACENG-----DVSRVKRLVDA--- 204
       ||||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| :
QY    121 HEMAHANQTASD-----SSVMVRQVGF 143
       ||||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| :
Db     205 -----ANTVNKDMAGRKSSPLHFAAGFG 227

```

```

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65
66
67
68
69
70
71
72
73
74
75
76
77
78
79
80
81
82
83
84
85
86
87
88
89
90
91
92
93
94
95
96
97
98
99
100
101
102
103
104
105
106
107
108
109
110
111
112
113
114
115
116
117
118
119
120
121
122
123
124
125
126
127
128
129
130
131
132
133
134
135
136
137
138
139
140
141
142
143
144
145
146
147
148
149
150
151
152
153
154
155
156
157
158
159
160
161
162
163
164
165
166
167
168
169
170
171
172
173
174
175
176
177
178
179
180
181
182
183
184
185
186
187
188
189
190
191
192
193
194
195
196
197
198
199
200
201
202
203
204
205
206
207
208
209
210
211
212
213
214
215
216
217
218
219
220
221
222
223
224
225
226
227
228
229
230
231
232
233
234
235
236
237
238
239
240
241
242
243
244
245
246
247
248
249
250
251
252
253
254
255
256
257
258
259
260
261
262
263
264
265
266
267
268
269
270
271
272
273
274
275
276
277
278
279
280
281
282
283
284
285
286
287
288
289
290
291
292
293
294
295
296
297
298
299
300
301
302
303
304
305
306
307
308
309
310
311
312
313
314
315
316
317
318
319
320
321
322
323
324
325
326
327
328
329
330
331
332
333
334
335
336
337
338
339
340
341
342
343
344
345
346
347
348
349
350
351
352
353
354
355
356
357
358
359
360
361
362
363
364
365
366
367
368
369
370
371
372
373
374
375
376
377
378
379
380
381
382
383
384
385
386
387
388
389
390
391
392
393
394
395
396
397
398
399
400
401
402
403
404
405
406
407
408
409
410
411
412
413
414
415
416
417
418
419
420
421
422
423
424
425
426
427
428
429
430
431
432
433
434
435
436
437
438
439
440
441
442
443
444
445
446
447
448
449
450
451
452
453
454
455
456
457
458
459
460
461
462
463
464
465
466
467
468
469
470
471
472
473
474
475
476
477
478
479
480
481
482
483
484
485
486
487
488
489
490
491
492
493
494
495
496
497
498
499
500
501
502
503
504
505
506
507
508
509
510
511
512
513
514
515
516
517
518
519
520
521
522
523
524
525
526
527
528
529
530
531
532
533
534
535
536
537
538
539
540
541
542
543
544
545
546
547
548
549
550
551
552
553
554
555
556
557
558
559
560
561
562
563
564
565
566
567
568
569
570
571
572
573
574
575
576
577
578
579
580
581
582
583
584
585
586
587
588
589
590
591
592
593
594
595
596
597
598
599
600
601
602
603
604
605
606
607
608
609
610
611
612
613
614
615
616
617
618
619
620
621
622
623
624
625
626
627
628
629
630
631
632
633
634
635
636
637
638
639
640
641
642
643
644
645
646
647
648
649
650
651
652
653
654
655
656
657
658
659
660
661
662
663
664
665
666
667
668
669
670
671
672
673
674
675
676
677
678
679
680
681
682
683
684
685
686
687
688
689
690
691
692
693
694
695
696
697
698
699
700
701
702
703
704
705
706
707
708
709
710
711
712
713
714
715
716
717
718
719
720
721
722
723
724
725
726
727
728
729
730
731
732
733
734
735
736
737
738
739
740
741
742
743
744
745
746
747
748
749
750
751
752
753
754
755
756
757
758
759
760
761
762
763
764
765
766
767
768
769
770
771
772
773
774
775
776
777
778
779
780
781
782
783
784
785
786
787
788
789
790
791
792
793
794
795
796
797
798
799
800
801
802
803
804
805
806
807
808
809
810
811
812
813
814
815
816
817
818
819
820
821
822
823
824
825
826
827
828
829
830
831
832
833
834
835
836
837
838
839
840
841
842
843
844
845
846
847
848
849
850
851
852
853
854
855
856
857
858
859
860
861
862
863
864
865
866
867
868
869
870
871
872
873
874
875
876
877
878
879
880
881
882
883
884
885
886
887
888
889
890
891
892
893
894
895
896
897
898
899
900
901
902
903
904
905
906
907
908
909
910
911
912
913
914
915
916
917
918
919
920
921
922
923
924
925
926
927
928
929
930
931
932
933
934
935
936
937
938
939
940
941
942
943
944
945
946
947
948
949
950
951
952
953
954
955
956
957
958
959
960
961
962
963
964
965
966
967
968
969
970
971
972
973
974
975
976
977
978
979
980
981
982
983
984
985
986
987
988
989
990
991
992
993
994
995
996
997
998
999

```

Query Match 11.7%; Score 90.5; DB 3; Length 1327;  
Best Local Similarity 28.4%; Pred. No. 2.8;

Matches	42;	Conservative	16;	Mismatches	61;	Indels	29;	Gaps	6;
QY	6	VAAFAA1-VVSGSALAGVWPOMGGGNNHGGNSGPDSTLSIYQYGSANAALQSDAR	64						
Db	99	VAAAPVVPVSTSAAGVAPNAGSGNNSPSSSSPTSS-SSSSPSPGSSLAESPAA	157						
QY	65	KSETTIT-----QSGYGNADVGGADNSTIETQTQGFNNATIDQWNAKNDQLVTRVVT	120						
Db	158	GVSSTAPLPGGAAGPTGVPVSGALRELLACRNG-----DVSrvKRLVDA---	204						
QY	121	HEMAHANOTASD-----SSVMVRQVGF	143						
Db	205	-----ANVNAKDWAGRKSSPLHFAAGF	227						

```

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65
66
67
68
69
70
71
72
73
74
75
76
77
78
79
80
81
82
83
84
85
86
87
88
89
90
91
92
93
94
95
96
97
98
99
100
101
102
103
104
105
106
107
108
109
110
111
112
113
114
115
116
117
118
119
120
121
122
123
124
125
126
127
128
129
130
131
132
133
134
135
136
137
138
139
140
141
142
143
144
145
146
147
148
149
150
151
152
153
154
155
156
157
158
159
160
161
162
163
164
165
166
167
168
169
170
171
172
173
174
175
176
177
178
179
180
181
182
183
184
185
186
187
188
189
190
191
192
193
194
195
196
197
198
199
200
201
202
203
204
205
206
207
208
209
210
211
212
213
214
215
216
217
218
219
220
221
222
223
224
225
226
227
228
229
230
231
232
233
234
235
236
237
238
239
240
241
242
243
244
245
246
247
248
249
250
251
252
253
254
255
256
257
258
259
260
261
262
263
264
265
266
267
268
269
270
271
272
273
274
275
276
277
278
279
280
281
282
283
284
285
286
287
288
289
290
291
292
293
294
295
296
297
298
299
300
301
302
303
304
305
306
307
308
309
310
311
312
313
314
315
316
317
318
319
320
321
322
323
324
325
326
327
328
329
330
331
332
333
334
335
336
337
338
339
340
341
342
343
344
345
346
347
348
349
350
351
352
353
354
355
356
357
358
359
360
361
362
363
364
365
366
367
368
369
370
371
372
373
374
375
376
377
378
379
380
381
382
383
384
385
386
387
388
389
390
391
392
393
394
395
396
397
398
399
400
401
402
403
404
405
406
407
408
409
410
411
412
413
414
415
416
417
418
419
420
421
422
423
424
425
426
427
428
429
430
431
432
433
434
435
436
437
438
439
440
441
442
443
444
445
446
447
448
449
450
451
452
453
454
455
456
457
458
459
460
461
462
463
464
465
466
467
468
469
470
471
472
473
474
475
476
477
478
479
480
481
482
483
484
485
486
487
488
489
490
491
492
493
494
495
496
497
498
499
500
501
502
503
504
505
506
507
508
509
510
511
512
513
514
515
516
517
518
519
520
521
522
523
524
525
526
527
528
529
530
531
532
533
534
535
536
537
538
539
540
541
542
543
544
545
546
547
548
549
550
551
552
553
554
555
556
557
558
559
560
561
562
563
564
565
566
567
568
569
570
571
572
573
574
575
576
577
578
579
580
581
582
583
584
585
586
587
588
589
590
591
592
593
594
595
596
597
598
599
600
601
602
603
604
605
606
607
608
609
610
611
612
613
614
615
616
617
618
619
620
621
622
623
624
625
626
627
628
629
630
631
632
633
634
635
636
637
638
639
640
641
642
643
644
645
646
647
648
649
650
651
652
653
654
655
656
657
658
659
660
661
662
663
664
665
666
667
668
669
670
671
672
673
674
675
676
677
678
679
680
681
682
683
684
685
686
687
688
689
690
691
692
693
694
695
696
697
698
699
700
701
702
703
704
705
706
707
708
709
710
711
712
713
714
715
716
717
718
719
720
721
722
723
724
725
726
727
728
729
730
731
732
733
734
735
736
737
738
739
740
741
742
743
744
745
746
747
748
749
750
751
752
753
754
755
756
757
758
759
760
761
762
763
764
765
766
767
768
769
770
771
772
773
774
775
776
777
778
779
780
781
782
783
784
785
786
787
788
789
790
791
792
793
794
795
796
797
798
799
800
801
802
803
804
805
806
807
808
809
810
811
812
813
814
815
816
817
818
819
820
821
822
823
824
825
826
827
828
829
830
831
832
833
834
835
836
837
838
839
840
841
842
843
844
845
846
847
848
849
850
851
852
853
854
855
856
857
858
859
860
861
862
863
864
865
866
867
868
869
870
871
872
873
874
875
876
877
878
879
880
881
882
883
884
885
886
887
888
889
890
891
892
893
894
895
896
897
898
899
900
901
902
903
904
905
906
907
908
909
910
911
912
913
914
915
916
917
918
919
920
921
922
923
924
925
926
927
928
929
930
931
932
933
934
935
936
937
938
939
940
941
942
943
944
945
946
947
948
949
950
951
952
953
954
955
956
957
958
959
960
961
962
963
964
965
966
967
968
969
970
971
972
973
974
975
976
977
978
979
980
981
982
983
984
985
986
987
988
989
990
991
992
993
994
995
996
997
998
999

```

	Query Match	11.7%;	Score 90.5;	DB 4;	Length 1327;
	Best Local Similarity	28.4%;	Prod. NO. 2.8;		
	Matches 42;	Conservative	16;	Mismatches 61;	Indels 29;
					Gaps 6;
QY	6	VAAFAAAL-VVSGSALACGWQWGGGNGHNGGNSGPDSTLTYIYOGSANAALALQSDAR	64		
Db	99	VAAAPVVPVAVSTSAAGVAEPNAPAGSGNNPSSSSSPTSS-SSSSPSPGGSIAESPEAA	157		
QY	65	KSEITIT-----QSGYGNAGADVGGADNSITETQTQNGFRNNAITDOWNAKNDQLVTRVVT	120		
Db	158	GVSSITAPLPGGAAGPGTGVPAVSGALRELLAEACNG-----DVSVRKRLVDA---	204		



QY 121 HEMAHANTASD-----SSVMVRQVGF 143  
Db 205 -----ANVNKDMAGRKSSPLHFAAGF 227

## RESULT 9

US-09-972-115A-8  
; Sequence 8, Application US/09972115A  
; Patent No. 6599728  
; GENERAL INFORMATION:  
; APPLICANT: Geron Corporation  
; APPLICANT: Gregg, Morin B.  
; APPLICANT: Walter, Funk D.  
; APPLICANT: Misczyk, Piatyszek A.  
; TITLE OF INVENTION: A Second Mammalian Telomerase  
; FILE REFERENCE: 080/003C  
; CURRENT APPLICATION NUMBER: US/09/972.115A  
; CURRENT FILING DATE: 2001-10-05  
; PRIOR APPLICATION NUMBER: US 60/128,577  
; PRIOR FILING DATE: 2000-04-10  
; PRIOR APPLICATION NUMBER: US 60/129,123  
; PRIOR FILING DATE: 1999-04-13  
; NUMBER OF SEQ ID NOS: 64  
; SOFTWARE: Patent in version 3.1  
; SEQ ID NO 8  
; LENGTH: 1327  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-09-972-115A-8

Query Match 11.7%; Score 90.5; DB 4; Length 1327;  
Best Local Similarity 28.4%; Pred. No. 2.8;  
Matches 42; Conservative 16; Mismatches 61; Indels 29; Gaps 6;

QY 6 VAAFAAI-VVSGSALAGVVPQWGGGNGHNGSGDPSTLSIYQGSANAALALQSDAR 64  
Db 99 VAAFPVPVNSTSGAAGVAENPAGSGNNPSSSSPTSS-SSSSSSPGSSLAESPAA 157  
QY 65 KSETTIT-----QSGYGNAGDVGGQADNSTIETQNGFRNNATIDONNAKYNPOLVTRVVT 120  
Db 158 GVSSTAPLPGAGGFG\*GVPAVSGALRELEACRNG-----DVSRRKRLVDA--- 204  
QY 121 HEMAHANTASD-----SSVMVRQVGF 143  
Db 205 -----ANVNKDMAGRKSSPLHFAAGF 227

## RESULT 10

US-08-864-038A-3  
; Sequence 3, Application US/08864038A  
; Patent No. 6001592  
; GENERAL INFORMATION:  
; APPLICANT: Kunio NAKASHIMA et al.

; TITLE OF INVENTION: NOVEL POLYPEPTIDE GENE CDNA, VECTOR  
; TITLE OF INVENTION: CONTAINING SAID CDNA, HOST CELLS TRANSFORMED WITH SAID  
; TITLE OF INVENTION: VECTOR, POLYPEPTIDE PRODUCED THEREBY, METHOD OF PRODUCING  
; TITLE OF INVENTION: SAID POLYPEPTIDE, DNA ENCODING SAID POLYPEPTIDE AND ANTIBODY  
; TITLE OF INVENTION: TO SAID POLYPEPTIDE  
; NUMBER OF SEQUENCES: 4  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: 812-5 Hirano  
; STREET: Isehinden  
; CITY: Tsu-city  
; STATE: Mie-prefecture  
; COUNTRY: JAPAN  
; ZIP: 514-01  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Diskette, 3.50 inch, 1.44 MB storage  
; COMPUTER: IBM Compatible  
; OPERATING SYSTEM: Microsoft Windows 95  
; SOFTWARE: Word Perfect 6.1  
; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/864,038A  
; FILING DATE: May 28, 1997  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: JP 8-184459  
; FILING DATE: 15-July-1996  
; ATTORNEY/AGENT INFORMATION:  
; NAME: C. Bruce Hamburg  
; REGISTRATION NUMBER: 22,389  
; REFERENCE/DOCKET NUMBER: F-5610  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (212)986-2340  
; TELEFAX: (212)953-7733  
; INFORMATION FOR SEQ ID NO: 3:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 738  
; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
; ORIGINAL SOURCE:  
; ORGANISM: Pinctada fucata  
; CELL TYPE: mantle epithelial cell  
; FEATURE:  
; NAME/KEY: peptide  
; LOCATION: from 1 to 738  
; IDENTIFICATION METHOD: E (by experiment)  
US-08-864-038A-3

Query Match 11.5%; Score 89.5; DB 3; Length 738;  
Best Local Similarity 35.4%; Pred. No. 1.6;  
Matches 29; Conservative 4; Mismatches 26; Indels 23; Gaps 3;

QY 3 LLKVAFAAI-VVSGSALAGVVPQWGGGNGHNGSGDPSTLSIYQGSANAALALQSD 62  
Db 419 LLKSSASASASASASAG-----GGGGGNGGGGGG-----GGGAGALA---- 460  
QY 63 ARKSETTITQSGYGNAGDVGG 84  
Db 461 -----AALAAAGAGGGGCGGG 477

## RESULT 11

US-09-198-452A-171  
; Sequence 171, Application US/09198452A  
; Patent No. 5559294  
; GENERAL INFORMATION:  
; APPLICANT: Griffiths, R.  
; TITLE OF INVENTION: Chlamydia pneumoniae genomic sequence and polypeptides, fragments  
; TITLE OF INVENTION: thereof and uses thereof, in particular for the diagnosis, prevention  
; TITLE OF INVENTION: and treatment of infection  
; FILE REFERENCE: 9710-003-999  
; CURRENT APPLICATION NUMBER: US/09/198,452A  
; CURRENT FILING DATE: 1998-11-24  
; NUMBER OF SEQ ID NOS: 6849  
; SEQ ID NO 171  
; LENGTH: 1156  
; TYPE: PRT  
; ORGANISM: Chlamydia pneumoniae  
US-09-198-452A-171

Query Match 11.2%; Score 87; DB 4; Length 1156;  
Best Local Similarity 32.0%; Pred. No. 5.1;  
Matches 32; Conservative 14; Mismatches 34; Indels 20; Gaps 5;

QY 44 TLSIYQGSANAALALQSDARKSETTITQSGYGNAGDVGGQADNSTI-----ELTQ 94  
Db 859 TVSPEDYAAVQAALA--AYVRKHESLIV-STYGLGAQEGQTSSKVTTLMRDLHAEELVE 915  
QY 95 NGFRNATIDQWNAKYNVDQLVTRVTHMAHANCTASDSS 134  
Db 916 MG-----VETRLNRSQILHRV--HSLHLHSLRSDSS 947

## RESULT 12

```

QY      126 ANQTASDSSVMVRQVGFENN 145
       ||| :||| :|||
Db      180 ANQATGESSTV---AGGSNN 196

RESULT 14
US-07-731-157A-2
; Sequence 2, Application US/07731157A
; Patent No. 5457032
; GENERAL INFORMATION:
; APPLICANT: Quax, Wilhelmus J.
; APPLICANT: Misset, Onno
; APPLICANT: Van der Laan, Jan M.
; APPLICANT: Lenting, Herman B.M.
; TITLE OF INVENTION: Mutated beta-lactam acylase genes
; NUMBER OF SEQUENCES: 50
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: COOLEY GODWARD CASTRO HUDDLESON & TATUM
; STREET: FIVE PALO ALTO SQUARE, 4TH FLOOR
; CITY: PALO ALTO
; STATE: CALIFORNIA
; COUNTRY: USA
; ZIP: 94306
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25 (EPO)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07731.157A
; FILING DATE: 19910509
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: EP 90200962
; FILING DATE: 18-APR-1990
; ATTORNEY/AGENT INFORMATION:
; NAME: RAE-VENTER PH.D.; BARBARA
; REGISTRATION NUMBER: 32,750
; REFERENCE/DOCKET NUMBER: GBRO-027/00U5
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415-494-7622
; TELEFAX: 415-857-0663
; TELEX: 380816 COOLEY PA
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 720 amino acids
; TYPE: AMINO ACID
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-07-731-157A-2

Query Match          10.5%; Score 81.5; DB 1; Length 720;
Best Local Similarity 25.8%; Pred. No. 9.5;
Matches 46; Conservative 17; Mismatches 56; Indels 59; Gaps 9

QY      1 KMLKVAFAAIIVVSGSALAGV-----PQ-----WGGGNGHNGG 36
       :||| :||| :|||
Db      2 LRVLHRAASALVMTATIGLPAFAFAEAETPTFPQAPIAYKPRSEIILWDG----YGVP 57

QY      37 NSSGPDSLTSLTYQVSANA-----ALALQSDARKSETTITQSQGYNGADV-GQGADNST 89
       :||| :||| :|||
Db      58 HIYGVDPSPAFYGVGWQAQRSGDNILRLYGAR-----CKGLAYWGPDYEQT 106

QY      90 IELTFONGFRNNAIDQWNAK-----NYDQLVTRVVTHEAHANQTASDSSVMVRQV 140
       :||| :||| :|||
Db      107 VWLLTNGVPERA--QQWYAQQSPDFRANLDAPAAGI---NAYAQQNPDISPDVRQV 158

RESULT 15
US-08-541-780-2
; Sequence 2, Application US/08541780
; Patent No. 5935831
; GENERAL INFORMATION:

```

Search completed: August 2, 2004, 14:58:35  
Job time : 12 secs

This Page Blank (uspto)

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: August 2, 2004, 14:54:48 ; Search time 36.8 Seconds  
(without alignments)  
1287.123 Million cell updates/sec

Title: US-09-543-407-24

Perfect score: 775  
Sequence: 1 MKLLKVAFAAIVVSGSALA.....DSSVMTRQVGGNNATANQY 151

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 1291235 seqs, 313682936 residues

Total number of hits satisfying chosen parameters: 1291235

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Published Applications AA:\*\*

1: /cgn2\_6/ptodata/2/pubpaa/US07\_PUBCOMB.pcp.\*  
2: /cgn2\_6/ptodata/2/pubpaa/PCT\_NEW\_PUB.pcp.\*  
3: /cgn2\_6/ptodata/2/pubpaa/US05\_NEW\_PUB.pcp.\*  
4: /cgn2\_6/ptodata/2/pubpaa/US06\_PUBCOMB.pcp.\*  
5: /cgn2\_6/ptodata/2/pubpaa/US07\_NEW\_PUB.pcp.\*  
6: /cgn2\_6/ptodata/2/pubpaa/PCTUS\_PUBCOMB.pcp.\*  
7: /cgn2\_6/ptodata/2/pubpaa/US08\_NEW\_PUB.pcp.\*  
8: /cgn2\_6/ptodata/2/pubpaa/US08\_PUBCOMB.pcp.\*  
9: /cgn2\_6/ptodata/2/pubpaa/US09A\_PUBCOMB.pcp.\*  
10: /cgn2\_6/ptodata/2/pubpaa/US09B\_PUBCOMB.pcp.\*  
11: /cgn2\_6/ptodata/2/pubpaa/US09C\_PUBCOMB.pcp.\*  
12: /cgn2\_6/ptodata/2/pubpaa/US09\_NEW\_PUB.pcp.\*  
13: /cgn2\_6/ptodata/2/pubpaa/US10A\_PUBCOMB.pcp.\*  
14: /cgn2\_6/ptodata/2/pubpaa/US10B\_PUBCOMB.pcp.\*  
15: /cgn2\_6/ptodata/2/pubpaa/US10C\_PUBCOMB.pcp.\*  
16: /cgn2\_6/ptodata/2/pubpaa/US10\_NEW\_PUB.pcp.\*  
17: /cgn2\_6/ptodata/2/pubpaa/US60\_NEW\_PUB.pcp.\*  
18: /cgn2\_6/ptodata/2/pubpaa/US60\_PUBCOMB.pcp.\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	530	68.4	151	12	US-09-741-873B-4
2	530	68.4	151	12	US-09-741-873B-4
3	452	58.3	131	12	US-09-741-873B-2
4	452	58.3	131	12	US-09-741-873B-2
5	111	14.3	445	15	US-10-369-493-20638
6	91	11.7	688	14	US-10-032-585-7876
7	90.5	11.7	263	12	US-10-425-114-49960
8	90.5	11.7	673	9	US-09-841-835-8
9	90.5	11.7	949	9	US-09-841-835-10
10	90.5	11.7	1327	9	US-09-841-835-2
11	90.5	11.7	1327	10	US-09-972-115A-8
12	90.5	11.7	1327	14	US-10-199-937-4
13	89	11.5	369	12	US-10-425-114-56041
14	89	11.5	486	12	US-10-424-599-275468
15	89	11.5	507	12	US-10-425-114-57763

16	88.5	11.4	145	16	US-10-437-963-147748	Sequence 147748,
17	88.5	11.4	438	14	US-10-156-761-9343	Sequence 9343, Ap
18	87	11.2	1156	15	US-10-289-762-171	Sequence 171, App
19	86	11.1	186	12	US-10-282-122A-49412	Sequence 49412, A
20	86	11.1	482	14	US-10-156-761-8763	Sequence 8763, Ap
21	86	11.1	762	12	US-10-282-122A-59874	Sequence 59874, A
22	85.5	11.0	270	16	US-10-437-963-122263	Sequence 122263,
23	85.5	11.0	271	14	US-10-156-761-11721	Sequence 11721, A
24	85.5	11.0	2457	12	US-10-282-122A-49854	Sequence 49854, A
25	85	11.0	478	16	US-10-437-963-115033	Sequence 115033,
26	84	10.8	254	10	US-09-880-748-1136	Sequence 1136, Ap
27	84	10.8	254	10	US-09-880-748-1165	Sequence 1165, Ap
28	84	10.8	254	12	US-10-293-418-1136	Sequence 1136, Ap
29	84	10.8	254	12	US-10-293-418-1165	Sequence 1165, Ap
30	83.5	10.8	350	14	US-10-156-761-13168	Sequence 13168, A
31	83.5	10.8	394	16	US-10-437-963-163379	Sequence 163379,
32	83	10.7	242	12	US-10-425-114-61520	Sequence 61520, A
33	83	10.7	889	10	US-09-952-267-15	Sequence 15, Appl
34	82.5	10.6	1074	12	US-10-282-122A-50616	Sequence 50616, A
35	82	10.6	214	16	US-10-437-963-119143	Sequence 119143,
36	82	10.6	254	10	US-09-880-748-1226	Sequence 1226, Ap
37	82	10.6	254	12	US-10-293-418-1226	Sequence 1226, Ap
38	82	10.6	678	12	US-10-282-122A-54573	Sequence 54573, A
39	82	10.6	883	12	US-10-282-122A-42626	Sequence 42626, A
40	81.5	10.5	189	16	US-10-437-963-170736	Sequence 170736,
41	81.5	10.5	256	14	US-10-156-761-11972	Sequence 11972, A
42	81.5	10.5	429	16	US-10-437-963-124678	Sequence 124678,
43	81.5	10.5	449	12	US-10-424-599-285485	Sequence 285485,
44	81.5	10.5	509	12	US-10-425-114-48393	Sequence 48393, A
45	81.5	10.5	529	12	US-10-425-114-49406	Sequence 49406, A

#### ALIGNMENTS

##### RESULT 1

US-09-741-873B-4  
; Sequence 4, Application US/09741873B  
; Publication No. US20020081722A1  
; GENERAL INFORMATION:  
; APPLICANT: Normark, Staffan  
; APPLICANT: Olsen, Arne  
; TITLE OF INVENTION: Fibronectin Binding Protein As Well As Its Preparation  
; FILE REFERENCE: 012889-084  
; CURRENT APPLICATION NUMBER: US/09/741,873B  
; CURRENT FILING DATE: 2003-04-04  
; PRIOR APPLICATION NUMBER: SE 8801723-1  
; PRIOR FILING DATE: 1998-05-06  
; PRIOR APPLICATION NUMBER: US 08/978,878  
; PRIOR FILING DATE: 1997-11-26  
; PRIOR APPLICATION NUMBER: US 07/347,189  
; PRIOR FILING DATE: 1989-05-04  
; PRIOR APPLICATION NUMBER: US 07/789,437  
; PRIOR FILING DATE: 1991-11-06  
; PRIOR APPLICATION NUMBER: US 07/970,846  
; PRIOR FILING DATE: 1992-11-03  
; PRIOR APPLICATION NUMBER: US 08/187,865  
; PRIOR FILING DATE: 1994-01-28  
; PRIOR APPLICATION NUMBER: US 08/318,519  
; PRIOR FILING DATE: 1994-10-05  
; NUMBER OF SEQ ID NOS: 10  
; SOFTWARE: Patent in version 3.0  
; SEQ ID NO 4  
; LENGTH: 151  
; TYPE: PRT  
; ORGANISM: Escherichia coli  
US-09-741-873B-4

Query Match 68.4%; Score 530; DB 12; Length 151;  
Best Local Similarity 69.5%; Pred. No. 7.2e-46;  
Matches 105; Conservative 19; Mismatches 27; Indels 0; Gaps 0;

QY 1 MKLLKVAFAAIVVSGSALAGWPGWGGGNNHNGSGSDSTLSIYQYGSANALALQ 60

Db 1 MKLLKVAALAAVFGSSAVGVPPQYGGGNGHGGGNSGPNSELNIYYQGGNSALALQ 60  
QY 61 SPARKSETTITQSGYNGADVCGGADNSTIELTQNGFRNATIDOWNAKNYDQLVTRVVT 120  
Db 61 TDARNSDLITTHGGGNGADVCGGSDSSIDLTORFGNSATLDQNGKNSMTVKQFGG 120  
QY 121 HEMAHANOTASDSSVMVROVGFNNATANQY 151  
Db 121 GNGAAVDQTASNSVNVTVQVGFNNATAHQY 151

RESULT 2  
US-09-741-873B-4  
; Sequence 4, Application US/09741873B  
; Publication No. US20040096965A9  
; GENERAL INFORMATION:  
; APPLICANT: Normark, Staffan  
; APPLICANT: Olsen, Arne  
; TITLE OF INVENTION: Fibronectin Binding Protein As Well As Its Preparation  
; FILE REFERENCE: 012889-084  
; CURRENT APPLICATION NUMBER: US/09/741,873B  
; CURRENT FILING DATE: 2003-04-04  
; PRIOR FILING DATE: 1998-05-06  
; PRIOR APPLICATION NUMBER: US 08/978,878  
; PRIOR FILING DATE: 1997-11-26  
; PRIOR APPLICATION NUMBER: US 07/347,189  
; PRIOR FILING DATE: 1989-05-04  
; PRIOR APPLICATION NUMBER: US 07/789,437  
; PRIOR FILING DATE: 1991-11-06  
; PRIOR APPLICATION NUMBER: US 07/970,846  
; PRIOR FILING DATE: 1992-11-03  
; PRIOR APPLICATION NUMBER: US 08/187,865  
; PRIOR FILING DATE: 1994-01-28  
; PRIOR APPLICATION NUMBER: US 08/318,519  
; NUMBER OF SEQ ID NOS: 10  
; SOFTWARE: Patent in version 3.0  
; SEQ ID NO 4  
; LENGTH: 151  
; TYPE: PRT  
; ORGANISM: Escherichia coli  
US-09-741-873B-4

Query Match 68.4%; Score 530; DB 12; Length 151;  
Best Local Similarity 69.5%; Pred. No. 7.2e-46;  
Matches 105; Conservative 19; Mismatches 27; Indels 0; Gaps 0;  
QY 1 MKLLKVAALAAVFGSSALAGVPPQYGGGNGHGGGNSGPNSELNIYYQGSANAALALQ 60  
Db 1 MKLLKVAALAAVFGSSAVGVPPQYGGGNGHGGGNSGPNSELNIYYQGGNSALALQ 60  
QY 61 SPARKSETTITQSGYNGADVCGGADNSTIELTQNGFRNATIDOWNAKNYDQLVTRVVT 120  
Db 61 TDARNSDLITTHGGGNGADVCGGSDSSIDLTORFGNSATLDQNGKNSMTVKQFGG 120  
QY 121 HEMAHANOTASDSSVMVROVGFNNATANQY 151  
Db 121 GNGAAVDQTASNSVNVTVQVGFNNATAHQY 151

RESULT 3  
US-09-741-873B-2  
; Sequence 2, Application US/09741873B  
; Publication No. US20020081722A1  
; GENERAL INFORMATION:  
; APPLICANT: Normark, Staffan  
; APPLICANT: Olsen, Arne  
; TITLE OF INVENTION: Fibronectin Binding Protein As Well As Its Preparation  
; FILE REFERENCE: 012889-084  
; CURRENT APPLICATION NUMBER: US/09/741,873B  
; CURRENT FILING DATE: 2003-04-04

; PRIOR APPLICATION NUMBER: SE 8801723-1  
; PRIOR FILING DATE: 1998-05-06  
; PRIOR APPLICATION NUMBER: US 08/978,878  
; PRIOR FILING DATE: 1997-11-26  
; PRIOR APPLICATION NUMBER: US 07/347,189  
; PRIOR FILING DATE: 1989-05-04  
; PRIOR APPLICATION NUMBER: US 07/789,437  
; PRIOR FILING DATE: 1991-11-06  
; PRIOR APPLICATION NUMBER: US 07/970,846  
; PRIOR FILING DATE: 1992-11-03  
; PRIOR APPLICATION NUMBER: US 08/187,865  
; PRIOR FILING DATE: 1994-01-28  
; PRIOR APPLICATION NUMBER: US 08/318,519  
; PRIOR FILING DATE: 1994-10-05  
; NUMBER OF SEQ ID NOS: 10  
; SOFTWARE: Patent in version 3.0  
; SEQ ID NO 2  
; LENGTH: 131  
; TYPE: PRT  
; ORGANISM: Escherichia coli  
US-09-741-873B-2

Query Match 58.3%; Score 452; DB 12; Length 131;  
Best Local Similarity 66.4%; Pred. No. 5e-38;  
Matches 87; Conservative 18; Mismatches 26; Indels 0; Gaps 0;  
QY 21 GVPYQWGGGNGHGGGNSGPNSELNIYYQGSANAALALQSDARKSETTITQSGYNGAD 80  
Db 1 GVPYQWGGGNGHGGGNSGPNSELNIYYQGGNSALALQTDARNSDLITTHGGGNGAD 60  
QY 81 VGGADNSTIELTQNGFRNATIDOWNAKNYDQLVTRVVTHEMAHANOTASDSSVMVROV 140  
Db 61 VGGSDSSIDLTORFGNSATLDQNGKNSMTVKQFGGNGAAVDQTASNSVNVTVQV 120  
QY 141 GFGNNATANQY 151  
Db 121 GFGNNATAHQY 131  
RESULT 4  
US-09-741-873B-2  
; Sequence 2, Application US/09741873B  
; Publication No. US20040096965A9  
; GENERAL INFORMATION:  
; APPLICANT: Normark, Staffan  
; APPLICANT: Olsen, Arne  
; TITLE OF INVENTION: Fibronectin Binding Protein As Well As Its Preparation  
; FILE REFERENCE: 012889-084  
; CURRENT APPLICATION NUMBER: US/09/741,873B  
; CURRENT FILING DATE: 2003-04-04  
; PRIOR APPLICATION NUMBER: SE 8801723-1  
; PRIOR FILING DATE: 1998-05-06  
; PRIOR APPLICATION NUMBER: US 08/978,878  
; PRIOR FILING DATE: 1997-11-26  
; PRIOR APPLICATION NUMBER: US 07/347,189  
; PRIOR FILING DATE: 1989-05-04  
; PRIOR APPLICATION NUMBER: US 07/789,437  
; PRIOR FILING DATE: 1991-11-06  
; PRIOR APPLICATION NUMBER: US 07/970,846  
; PRIOR FILING DATE: 1992-11-03  
; PRIOR APPLICATION NUMBER: US 08/187,865  
; PRIOR FILING DATE: 1994-01-28  
; PRIOR APPLICATION NUMBER: US 08/318,519  
; PRIOR FILING DATE: 1994-10-05  
; NUMBER OF SEQ ID NOS: 10  
; SOFTWARE: Patent in version 3.0  
; SEQ ID NO 2  
; LENGTH: 131  
; TYPE: PRT  
; ORGANISM: Escherichia coli  
US-09-741-873B-2

Best Local Similarity 66.4%; Pred. No. 5e-38;  
Matches 87; Conservative 18; Mismatches 26; Indels 0; Gaps 0;  
QY 21 GVPVPGGGGNNHGGSSGPDSTLSIYQYGSANAALALQSDARKSETTITQSGYNGAD 80  
Db 1 GVPVPGGGGNNHGGSSGPDSTLSIYQYGSANAALALQSDARKSETTITQSGYNGAD 60  
QY 81 VQGGADNSTIETQNGFRNNATIDQNAKNDYDQVTRVVTHEMAHANQATSDSSVMVROV 140  
Db 61 VQGGADNSTIETQNGFRNNATIDQNAKNDYDQVTRVVTHEMAHANQATSDSSVMVROV 120  
QY 141 GFGNNATAHQY 151  
Db 121 GFGNNATAHQY 131  
RESULT 5  
US-10-369-493-20638  
; Sequence 20638, Application US/10369493  
; Publication No. US20030233675A1  
; GENERAL INFORMATION:  
; APPLICANT: Cao, Yongwei  
; APPLICANT: Hinkle, Gregory J.  
; APPLICANT: Slater, Steven C.  
; APPLICANT: Goldman, Barry S.  
; APPLICANT: Chen, Xianfeng  
; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF  
; FILE REFERENCE: 38-10(52052)B  
; CURRENT APPLICATION NUMBER: US/10/369,493  
; PRIOR FILING DATE: 2003-02-28  
; PRIOR APPLICATION NUMBER: US 60/360,039  
; PRIOR FILING DATE: 2002-02-21  
; NUMBER OF SEQ ID NOS: 47374  
; SEQ ID NO 20638  
; LENGTH: 445  
; TYPE: PRT  
; ORGANISM: Rhodospseudomonas palustris  
; NAME/KEY: unsure  
; LOCATION: (1) (445)  
; OTHER INFORMATION: unsure at all Xaa locations  
US-10-369-493-20638

Query Match 14.3%; Score 111; DB 15; Length 445;  
Best Local Similarity 25.5%; Pred. No. 0.01;  
Matches 42; Conservative 25; Mismatches 68; Indels 30; Gaps 6;  
QY 7 AAFAA-----IVVGSALAGVVPQGGG-----NHNGG-----GNSSGPDSTLSIYQY 50  
Db 19 AAFADSNVTYLNQTCNQDQANITQSGNGSVGAFNGSGFLQENGTLTSGA-NLLITVKOS 77  
QY 51 GSANAALALQSDARKSETTITQSGYNGADVQGGADNSTIETQNGFRNNATIDQNAKND 110  
Db 78 GNSNSV-----GRDIQKQSGAGNSAAIFQEGTSDVQLQQTGTSNGAVPVGNNWTN 129  
QY 111 YDQVTRVVTHEMAHAN-----QTASDSSVMVROVQFGNNATNQ 150  
Db 130 DPGVFNKITQDSSSSNGSKSVIQDQGNVFSIKQGTGNSTSYNQ 174

RESULT 6-  
US-10-032-585-7876  
; Sequence 7876, Application US/10032585  
; Publication No. US20030180953A1  
; GENERAL INFORMATION:  
; APPLICANT: Terry, Roemer D.  
; APPLICANT: Bo, Jiang  
; APPLICANT: Charles, Boone  
; APPLICANT: Howard, Bussey  
; TITLE OF INVENTION: Gene Disruption Methodologies for Drug Target Discovery  
; FILE REFERENCE: 10182-005-999  
; CURRENT APPLICATION NUMBER: US/10/032,585

; CURRENT FILING DATE: 2001-12-20  
; NUMBER OF SEQ ID NOS: 8000  
; SOFTWARE: Patent in version 3.1  
; SEQ ID NO 7876  
; LENGTH: 688  
; TYPE: PRT  
; ORGANISM: Candida albicans  
US-10-032-585-7876  
Query Match 11.7%; Score 91; DB 14; Length 688;  
Best Local Similarity 23.7%; Pred. No. 1.9; Indels 6; Gaps 2;  
Matches 28; Conservative 25; Mismatches 59; Indels 21; Gaps 4;  
QY 30 GNHGGCGSSGPDSTLSIYQYGSANAALALQSDARKSETTITQSGYNGADVQGGADNST 89  
Db 488 GNHGGCGSSGPDSTLSIYQYGSANAALALQSDARKSETTITQSGYNGADVQGGADNST 544  
QY 90 IELTQNGFRNNATIDQNAKNDYDQVTRVVTHEMAHANQATSDSSVMVROVQFGNNAT 147  
Db 545 KSKKSNFDDNNSNS---ALNNLDKSKLINTNITNITSSSSSPVINHGGRSS 599  
RESULT 7  
US-10-425-114-49960  
; Sequence 49960, Application US/10425114  
; Publication No. US20040034888A1  
; GENERAL INFORMATION:  
; APPLICANT: Liu, Jingdong  
; APPLICANT: Zhou, Yihua  
; APPLICANT: Kovalic, David K.  
; APPLICANT: Screen, Steven E.  
; APPLICANT: Tabaska, Jack E.  
; APPLICANT: Cao, Yongwei  
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With  
; FILE REFERENCE: 38-21(53313)B  
; CURRENT APPLICATION NUMBER: US/10/425,114  
; CURRENT FILING DATE: 2003-04-28  
; NUMBER OF SEQ ID NOS: 73128  
; SEQ ID NO 49960  
; LENGTH: 263  
; TYPE: PRT  
; ORGANISM: Zea mays  
; FEATURE:  
; OTHER INFORMATION: Clone ID: 700071884\_FLI.pep  
US-10-425-114-49960  
Query Match 11.7%; Score 90.5; DB 12; Length 263;  
Best Local Similarity 27.9%; Pred. No. 0.62;  
Matches 29; Conservative 15; Mismatches 39; Indels 21; Gaps 4;  
QY 17 SALAGVVPQGGGNNHGGSSGPDSTLSIYQYGSANAALALQSDARKSETTITQSG-- 74  
Db 84 SSIAQ-----GGGGGQGGGGTNGSGSGGSGGSGGSGGSGGSGGSGGSGGSGGSGG 138  
QY 75 -----YNGGA--DVQGGADNSTIEL--TQNGFRNNATID 104  
Db 139 GGMGGGANGAYGSGAGGCGVKGEGVGVGVALAPSSNGYNGGAAD 182

RESULT 8  
US-09-841-835-8  
; Sequence 8, Application US/09841835  
; Patent No. US20020076795A1  
; GENERAL INFORMATION:  
; APPLICANT: de Lange, Titia  
; APPLICANT: Smith, Susan  
; TITLE OF INVENTION: A PROTEIN THAT BINDS TO TRF1 AND METHODS  
; TITLE OF INVENTION: OF USE THEREOF  
; NUMBER OF SEQUENCES: 12  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Klauber & Jackson  
; STREET: 411 Hackensack Avenue, 4th Floor

CITY: Hackensack  
STATE: New Jersey  
COUNTRY: USA  
ZIP: 07601  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent in Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/841,835  
FILING DATE:  
CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 09/196,387  
FILING DATE:  
ATTORNEY/AGENT INFORMATION:  
NAME: Jackson Esq., David A.  
REGISTRATION NUMBER: 26,742  
REFERENCE/DOCKET NUMBER: 600-1-230 CIP1  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 201-487-5800  
TELEFAX: 201-343-1684  
TELEX: 133521  
INFORMATION FOR SEQ ID NO: 10:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 949 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-09-841-835-10

Query Match 11.7%; Score 90.5; DB 9; Length 949;  
Best Local Similarity 28.4%; Pred. No. 3.1;  
Matches 42; Conservative 16; Mismatches 61; Indels 29; Gaps 6;

QY 6 VAAFAAI-VVGSALAGVVPQWGGGNNHNGGNSGPDSTLSIYQYGSANAALQSDAR 64  
DB 99 VAAAPVPAVSTSSAAGVAPNPAGSGNNSPSSSSPTSS-SSSPSPGSSLAESPEAA 157  
QY 65 KSETTIT-OSQYNGADVGCGADNNTIELTQNGFRNNAIDQWNAKNDQLVTRVVT 120  
DB 158 GVSSTAPLPGGAAGTGVPAVSGALRELEACRNG-----DVSRRKLVDA--- 204  
QY 121 HEMAHANQTASD-----SSVMVRQVGF 143  
DB 205 -----ANVNAKDMAGRKSSPLHFAAGFG 227

RESULT 10  
US-09-841-835-2  
Sequence 2, Application US/09841835  
Patent No. US20020076795A1  
GENERAL INFORMATION:  
APPLICANT: de Lange, Titia  
APPLICANT: Smith, Susan  
TITLE OF INVENTION: A PROTEIN THAT BINDS TO TRF1 AND METHODS  
TITLE OF INVENTION: OF USE THEREOF  
NUMBER OF SEQUENCES: 12  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Klauber & Jackson  
STREET: 411 Hackensack Avenue, 4th Floor  
CITY: Hackensack  
STATE: New Jersey  
COUNTRY: USA  
ZIP: 07601  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent in Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/841,835  
FILING DATE:  
CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 09/196,387  
FILING DATE:  
ATTORNEY/AGENT INFORMATION:  
NAME: Jackson Esq., David A.  
REGISTRATION NUMBER: 26,742  
REFERENCE/DOCKET NUMBER: 600-1-230 CIP1  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 201-487-5800

CITY: Hackensack  
STATE: New Jersey  
COUNTRY: USA  
ZIP: 07601  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent in Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/841,835  
FILING DATE:  
CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 09/196,387  
FILING DATE:  
ATTORNEY/AGENT INFORMATION:  
NAME: Jackson Esq., David A.  
REGISTRATION NUMBER: 26,742  
REFERENCE/DOCKET NUMBER: 600-1-230 CIP1  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 201-487-5800

Query Match 11.7%; Score 90.5; DB 9; Length 673;  
Best Local Similarity 28.4%; Pred. No. 2;  
Matches 42; Conservative 16; Mismatches 61; Indels 29; Gaps 6;

QY 6 VAAFAAI-VVGSALAGVVPQWGGGNNHNGGNSGPDSTLSIYQYGSANAALQSDAR 64  
DB 99 VAAAPVPAVSTSSAAGVAPNPAGSGNNSPSSSSPTSS-SSSPSPGSSLAESPEAA 157  
QY 65 KSETTIT-OSQYNGADVGCGADNNTIELTQNGFRNNAIDQWNAKNDQLVTRVVT 120  
DB 158 GVSSTAPLPGGAAGTGVPAVSGALRELEACRNG-----DVSRRKLVDA--- 204  
QY 121 HEMAHANQTASD-----SSVMVRQVGF 143  
DB 205 -----ANVNAKDMAGRKSSPLHFAAGFG 227

RESULT 9  
US-09-841-835-10  
Sequence 10, Application US/09841835  
Patent No. US20020076795A1  
GENERAL INFORMATION:  
APPLICANT: de Lange, Titia  
APPLICANT: Smith, Susan  
TITLE OF INVENTION: A PROTEIN THAT BINDS TO TRF1 AND METHODS  
TITLE OF INVENTION: OF USE THEREOF  
NUMBER OF SEQUENCES: 12  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Klauber & Jackson  
STREET: 411 Hackensack Avenue, 4th Floor  
CITY: Hackensack  
STATE: New Jersey  
COUNTRY: USA  
ZIP: 07601  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent in Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/841,835



TELEFAX: 201-343-1684  
TELEX: 133521  
INFORMATION FOR SEQ ID NO: 2:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1327 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
HYPOTHETICAL: NO  
US-09-841-835-2

Query Match  
Best Local Similarity 11.7%; Score 90.5; DB 9; Length 1327;  
Matches 42; Conservative 16; Mismatches 61; Indels 29; Gaps 6;

6 VAAFAAI-VVSGSALAGVVPQWGGGNGHNGSGPDSTLSIYQYGSANAALALQSDAR 64  
99 VAAAPVPAVSTSSAAGVAPNPAGSGNNSSPSSPTSS-SSSSPSPGSSLAESPEAA 157  
65 KSETTIT- - - - - QSGYNGADVQGGADNSTIELTQNGFRNNATIDOWNAKNYDQIVTRVVT 120  
158 GVSSTAPLPGGAAGPGTGVPAVSGALRELLEACRNG- - - - - DVSrvKRLVDA- - - 204

121 HEMAHANQTASD- - - - - SSVVMVRQVGF 143  
205 - - - - - ANVNADMAGRKSSPLHFAAGFG 227

RESULT 11  
US-09-972-115A-8  
Sequence 8, Application US/09972115A  
Publication No. US20030032769A1  
GENERAL INFORMATION:  
APPLICANT: Geron Corporation  
APPLICANT: Gregg, Morin B.  
APPLICANT: Walter, Funk D.  
APPLICANT: Mieczyslaw, Piatyszek A.  
TITLE OF INVENTION: A Second Mammalian Telomerase  
FILE REFERENCE: 080/003C  
CURRENT APPLICATION NUMBER: US/09/972,115A  
CURRENT FILING DATE: 2001-10-05  
PRIOR APPLICATION NUMBER: US 60/128,577  
PRIOR FILING DATE: 2000-04-10  
PRIOR APPLICATION NUMBER: US 60/129,123  
PRIOR FILING DATE: 1999-04-13  
NUMBER OF SEQ ID NOS: 64  
SOFTWARE: Patent in version 3.1  
SEQ ID NO 8  
LENGTH: 1327  
TYPE: PRT  
ORGANISM: Homo sapiens  
US-09-972-115A-8

Query Match  
Best Local Similarity 11.7%; Score 90.5; DB 10; Length 1327;  
Matches 42; Conservative 16; Mismatches 61; Indels 29; Gaps 6;

6 VAAFAAI-VVSGSALAGVVPQWGGGNGHNGSGPDSTLSIYQYGSANAALALQSDAR 64  
99 VAAAPVPAVSTSSAAGVAPNPAGSGNNSSPSSPTSS-SSSSPSPGSSLAESPEAA 157  
65 KSETTIT- - - - - QSGYNGADVQGGADNSTIELTQNGFRNNATIDOWNAKNYDQIVTRVVT 120  
158 GVSSTAPLPGGAAGPGTGVPAVSGALRELLEACRNG- - - - - DVSrvKRLVDA- - - 204

121 HEMAHANQTASD- - - - - SSVVMVRQVGF 143  
205 - - - - - ANVNADMAGRKSSPLHFAAGFG 227

RESULT 12  
US-10-199-937-4

Sequence 4, Application US/10199937  
Publication No. US20030190739A1  
GENERAL INFORMATION:  
APPLICANT: Christenson, Erik  
APPLICANT: Demaggio, Anthony J.  
APPLICANT: Goldman, Phyllis S.  
APPLICANT: McElligott, David L.  
TITLE OF INVENTION: TANKYRASE2 MATERIALS AND METHODS  
FILE REFERENCE: 27866/36559  
CURRENT APPLICATION NUMBER: US/10/199,937  
CURRENT FILING DATE: 2002-07-22  
PRIOR APPLICATION NUMBER: US/09/606,035  
PRIOR FILING DATE: 2000-06-28  
PRIOR APPLICATION NUMBER: 60/141,582  
PRIOR FILING DATE: 1999-06-29  
NUMBER OF SEQ ID NOS: 178  
SOFTWARE: Patent in Ver. 2.0  
SEQ ID NO 4  
LENGTH: 1327  
TYPE: PRT  
ORGANISM: Homo sapiens  
US-10-199-937-4

Query Match  
Best Local Similarity 11.7%; Score 90.5; DB 14; Length 1327;  
Matches 42; Conservative 16; Mismatches 61; Indels 29; Gaps 6;

6 VAAFAAI-VVSGSALAGVVPQWGGGNGHNGSGPDSTLSIYQYGSANAALALQSDAR 64  
99 VAAAPVPAVSTSSAAGVAPNPAGSGNNSSPSSPTSS-SSSSPSPGSSLAESPEAA 157  
65 KSETTIT- - - - - QSGYNGADVQGGADNSTIELTQNGFRNNATIDOWNAKNYDQIVTRVVT 120  
158 GVSSTAPLPGGAAGPGTGVPAVSGALRELLEACRNG- - - - - DVSrvKRLVDA- - - 204

121 HEMAHANQTASD- - - - - SSVVMVRQVGF 143  
205 - - - - - ANVNADMAGRKSSPLHFAAGFG 227

RESULT 13  
US-10-425-114-56041  
Sequence 56041, Application US/10425114  
Publication No. US20040034888A1  
GENERAL INFORMATION:  
APPLICANT: Liu, Jingdong  
APPLICANT: Zhou, Yihua  
APPLICANT: Kovalic, David K.  
APPLICANT: Screen, Steven E.  
APPLICANT: Tabaska, Jack E.  
APPLICANT: Cao, Yongwei  
TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With  
TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement  
FILE REFERENCE: 38-21(53313)B  
CURRENT APPLICATION NUMBER: US/10/425,114  
CURRENT FILING DATE: 2003-04-28  
NUMBER OF SEQ ID NOS: 73128  
SEQ ID NO 56041  
LENGTH: 369  
TYPE: PRT  
ORGANISM: Glycine max  
FEATURE:  
OTHER INFORMATION: Clone ID: 701205720\_FLI.pep  
US-10-425-114-56041

Query Match  
Best Local Similarity 11.5%; Score 89; DB 12; Length 369;  
Matches 36; Conservative 20; Mismatches 61; Indels 20; Gaps 6;

14 VSGSALAGVVPQWGGGNGHNGSGPDSTLSIYQYGSANAALALQSDARKSETTITQS 73  
13 VRSSASKG-KPPLSAGNNNGGWNWDD- - - - - GFGSARGG- - - - - ADLRNQTGDVR 62



GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: August 2, 2004, 14:48:33 ; Search time 167.9 Seconds  
(without alignments)  
877.809 Million cell updates/sec

Title: US-09-543-407-24

Perfect score: 775

Sequence: 1 MKLLKVAFAAIVVSGSALA.....DSSVMVRQVGFNNATANQY 151

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 6019581 seqs, 976053577 residues

Total number of hits satisfying chosen parameters: 6019581

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Pending Patents AA\_Main:\*

- 1: /cgn2\_6/ptodata/2/paa/pctus\_COMB.pep.\*
- 2: /cgn2\_6/ptodata/2/paa/US06\_COMB.pep.\*
- 3: /cgn2\_6/ptodata/2/paa/US07\_COMB.pep.\*
- 4: /cgn2\_6/ptodata/2/paa/US08\_COMB.pep.\*
- 5: /cgn2\_6/ptodata/2/paa/US081\_COMB.pep.\*
- 6: /cgn2\_6/ptodata/2/paa/US082\_COMB.pep.\*
- 7: /cgn2\_6/ptodata/2/paa/US083\_COMB.pep.\*
- 8: /cgn2\_6/ptodata/2/paa/US084\_COMB.pep.\*
- 9: /cgn2\_6/ptodata/2/paa/US085\_COMB.pep.\*
- 10: /cgn2\_6/ptodata/2/paa/US086\_COMB.pep.\*
- 11: /cgn2\_6/ptodata/2/paa/US087\_COMB.pep.\*
- 12: /cgn2\_6/ptodata/2/paa/US088\_COMB.pep.\*
- 13: /cgn2\_6/ptodata/2/paa/US089\_COMB.pep.\*
- 14: /cgn2\_6/ptodata/2/paa/US090\_COMB.pep.\*
- 15: /cgn2\_6/ptodata/2/paa/US091\_COMB.pep.\*
- 16: /cgn2\_6/ptodata/2/paa/US092\_COMB.pep.\*
- 17: /cgn2\_6/ptodata/2/paa/US093\_COMB.pep.\*
- 18: /cgn2\_6/ptodata/2/paa/US094\_COMB.pep.\*
- 19: /cgn2\_6/ptodata/2/paa/US095\_COMB.pep.\*
- 20: /cgn2\_6/ptodata/2/paa/US096\_COMB.pep.\*
- 21: /cgn2\_6/ptodata/2/paa/US097A\_COMB.pep.\*
- 22: /cgn2\_6/ptodata/2/paa/US097B\_COMB.pep.\*
- 23: /cgn2\_6/ptodata/2/paa/US098\_COMB.pep.\*
- 24: /cgn2\_6/ptodata/2/paa/US099A\_COMB.pep.\*
- 25: /cgn2\_6/ptodata/2/paa/US099B\_COMB.pep.\*
- 26: /cgn2\_6/ptodata/2/paa/US100\_COMB.pep.\*
- 27: /cgn2\_6/ptodata/2/paa/US101\_COMB.pep.\*
- 28: /cgn2\_6/ptodata/2/paa/US102\_COMB.pep.\*
- 29: /cgn2\_6/ptodata/2/paa/US103\_COMB.pep.\*
- 30: /cgn2\_6/ptodata/2/paa/US104\_COMB.pep.\*
- 31: /cgn2\_6/ptodata/2/paa/US106\_COMB.pep.\*
- 32: /cgn2\_6/ptodata/2/paa/US107\_COMB.pep.\*
- 33: /cgn2\_6/ptodata/2/paa/US60\_COMB.pep.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
------------	-------	-------------	--------	----	-------------

1	775	100.0	151	19	US-09-543-407-24	Sequence 24, Appl
2	714	92.1	151	19	US-09-543-407-14	Sequence 14, Appl
3	700	90.3	151	19	US-09-543-407-30	Sequence 30, Appl
4	690	89.0	151	19	US-09-543-407-5	Sequence 5, Appl
5	689	88.9	151	6	US-08-233-642A-57	Sequence 57, Appl
6	655	84.5	151	19	US-09-543-407-12	Sequence 12, Appl
7	609	78.6	151	19	US-09-543-407-22	Sequence 22, Appl
8	608	78.5	151	19	US-09-543-407-18	Sequence 18, Appl
9	606	78.2	151	19	US-09-543-407-20	Sequence 20, Appl
10	603	77.8	151	19	US-09-543-407-31	Sequence 31, Appl
11	600	77.4	151	19	US-09-543-407-28	Sequence 28, Appl
12	575	74.2	151	19	US-09-543-407-16	Sequence 16, Appl
13	533	68.8	151	19	US-09-543-407-7	Sequence 7, Appl
14	530	68.4	151	13	US-08-978-878-4	Sequence 4, Appl
15	530	68.4	151	21	US-09-741-873B-4	Sequence 4, Appl
16	528	68.1	151	33	US-60-352-946-2	Sequence 2, Appl
17	528	68.1	151	33	US-60-444-371-2	Sequence 2, Appl
18	505	65.2	120	6	US-08-233-642A-55	Sequence 55, Appl
19	475	61.3	158	16	US-09-252-691-5834	Sequence 5834, Ap
20	475	61.3	158	16	US-09-252-691C-5834	Sequence 5834, Ap
21	475	61.3	158	30	US-10-417-886-5834	Sequence 34, Appl
22	471	60.8	109	19	US-09-543-407-34	Sequence 2, Appl
23	452	58.3	131	13	US-08-978-878-2	Sequence 2, Appl
24	452	58.3	131	21	US-09-741-873B-2	Sequence 35, Appl
25	348	44.9	109	19	US-09-543-407-35	Sequence 37, Appl
26	347	44.8	68	19	US-09-543-407-32	Sequence 32, Appl
27	248.5	32.1	70	19	US-09-543-407-39	Sequence 39, Appl
28	237	30.6	48	19	US-09-543-407-39	Sequence 20638, A
29	111	14.3	445	29	US-10-369-493-20638	Sequence 20638, A
30	111	14.3	445	33	US-60-360-039-20638	Sequence 5833, Ap
31	107	13.8	186	16	US-09-252-691-5833	Sequence 5833, Ap
32	107	13.8	186	16	US-09-252-691C-5833	Sequence 5833, Ap
33	107	13.8	186	30	US-10-417-886-5833	Sequence 5833, Ap
34	98.5	12.7	151	19	US-09-543-407-6	Sequence 6, Appl
35	98.5	12.7	151	19	US-09-543-407-8	Sequence 8, Appl
36	96.5	12.5	151	19	US-09-543-407-8	Sequence 44, Appl
37	92.5	11.9	850	20	US-09-614-150A-24084	Sequence 24084, A
38	92.5	11.9	850	20	US-09-614-150A-24084	Sequence 24179, A
39	92.5	11.9	850	33	US-60-191-681-19019	Sequence 19019, A
40	92.5	11.9	850	33	US-60-191-681-19019	Sequence 14916, A
41	92.5	11.9	1028	20	US-09-614-150A-14916	Sequence 14916, A
42	92.5	11.9	1028	20	US-09-614-150A-14916	Sequence 14978, A
43	92.5	11.9	1028	33	US-60-167-217-14978	Sequence 12194, A
44	92.5	11.9	1028	33	US-60-173-464-12194	
45	92.5	11.9	1028	33	US-60-173-464-12194	

ALIGNMENTS

RESULT 1  
US-09-543-407-24  
; Sequence 24, Application US/09543407  
; GENERAL INFORMATION:  
; APPLICANT: White, Aaron P.  
; APPLICANT: Doran, James L.  
; APPLICANT: Collinson, S. Karen  
; APPLICANT: Kay, William W.  
; TITLE OF INVENTION: BACTERIAL FIMBRIAL SYSTEM FOR PRESENTATION OF HETEROLOGOUS PEPTIDE SEQUENCES  
; TITLE OF INVENTION: PRESENTATION OF HETEROLOGOUS PEPTIDE SEQUENCES  
; FILE REFERENCE: 920043.406  
; CURRENT APPLICATION NUMBER: US/09/543,407  
; CURRENT FILING DATE: 2000-04-05  
; NUMBER OF SEQ ID NOS: 59  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 24  
; TYPE: PRT  
; LENGTH: 151  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: Recombinant Salmonella enteritidis 3b afga  
; OTHER INFORMATION: sequence containing the replacement fragment  
; OTHER INFORMATION: encoding PT3 from Gp63 of Leishmania major.

US-09-543-407-24

Query Match 100.0%; Score 775; DB 19; Length 151;  
Best Local Similarity 100.0%; Pred. No. 2.7e-74;  
Matches 151; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 MKLLKVAFAAIVVSGSALAGVVPQGGGNGHNGGNSGPDSTLSIYQYGSANAALALQ 60  
Db 1 MKLLKVAFAAIVVSGSALAGVVPQGGGNGHNGGNSGPDSTLSIYQYGSANAALALQ 60  
Qy 61 SDARKSETTITQSGYNGADVGQGDNSTIETQNGFRNNATIDQWNAKNDQLVTRVVT 120  
Db 61 SDARKSETTITQSGYNGADVGQGDNSTIETQNGFRNNATIDQWNAKNDQLVTRVVT 120  
Qy 121 HEMAHANQASDSSVMVRQVGFNNATANQY 151  
Db 121 HEMAHANQASDSSVMVRQVGFNNATANQY 151

RESULT 2

US-09-543-407-14

; Sequence 14, Application US/09543407  
; GENERAL INFORMATION:  
; APPLICANT: White, Aaron P.  
; APPLICANT: Doran, James L.  
; APPLICANT: Collinson, S. Karen  
; APPLICANT: Kay, William W.  
; TITLE OF INVENTION: BACTERIAL FIMBRIAL SYSTEM FOR  
; TITLE OF INVENTION: PRESENTATION OF HETEROLOGOUS PEPTIDE SEQUENCES  
; FILE REFERENCE: 920043.406  
; CURRENT APPLICATION NUMBER: US/09/543,407  
; CURRENT FILING DATE: 2000-04-05  
; NUMBER OF SEQ ID NOS: 59  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 14  
; LENGTH: 151  
; TYPE: PRT  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: Recombinant Salmonella enteritidis 3b afga  
; OTHER INFORMATION: sequence containing the replacement fragment  
; OTHER INFORMATION: encoding PT3 from GP63 of Leishmania major.  
US-09-543-407-14

Query Match 92.1%; Score 714; DB 19; Length 151;  
Best Local Similarity 91.1%; Pred. No. 9.2e-68;  
Matches 144; Conservative 0; Mismatches 0; Indels 14; Gaps 2;

Qy 1 MKLLKVAFAAIVVSGSALAGVVPQGGGNGHNGGNSGPDSTLSIYQYGSANAALALQ 60  
Db 1 MKLLKVAFAAIVVSGSALAGVVPQGGGNGHNGGNSGPDSTLSIYQYGSANAALALQ 60  
Qy 61 SDARKSETTITQSGYNGADVGQGDNSTIETQNGFRNNATIDQWNAKNDQLVTRVVT 113  
Db 61 SDARKSETTITQSGYNGADVGQGDNSTIETQNGFRNNATIDQWNAKNDQLVTRVVT 120  
Qy 114 LVTRVVTHEMANQASDSSVMVRQVGFNNATANQY 151  
Db 121 LVTRVVTHEMAHA-----SVVRQVGFNNATANQY 151

RESULT 3

US-09-543-407-30

; Sequence 30, Application US/09543407  
; GENERAL INFORMATION:  
; APPLICANT: White, Aaron P.  
; APPLICANT: Doran, James L.  
; APPLICANT: Collinson, S. Karen  
; APPLICANT: Kay, William W.  
; TITLE OF INVENTION: BACTERIAL FIMBRIAL SYSTEM FOR  
; TITLE OF INVENTION: PRESENTATION OF HETEROLOGOUS PEPTIDE SEQUENCES  
; FILE REFERENCE: 920043.406  
; CURRENT APPLICATION NUMBER: US/09/543,407

CURRENT FILING DATE: 2000-04-05

; NUMBER OF SEQ ID NOS: 59  
; SOFTWARE: FastSeq for Windows Version 4.0

SEQ ID NO 30

LENGTH: 151

TYPE: PRT

ORGANISM: Artificial Sequence

FEATURE:

; OTHER INFORMATION: Recombinant Salmonella enteritidis 3b afga  
; OTHER INFORMATION: sequence containing the replacement fragment

; OTHER INFORMATION: encoding PT3 from GP63 of Leishmania major.

US-09-543-407-30

Query Match 90.3%; Score 700; DB 19; Length 151;

Best Local Similarity 89.9%; Pred. No. 2.9e-66;

Matches 143; Conservative 0; Mismatches 0; Indels 16; Gaps 2;

Qy 1 MKLLKVAFAAIVVSGSALAGVVPQGGGNGHNGGNSGPDSTLSIYQYGSANAALALQ 60  
Db 1 MKLLKVAFAAIVVSGSALAGVVPQGGGNGHNGGNSGPDSTLSIYQYGSANAALALQ 60  
Qy 61 SDARKSETTITQSGYNGADVGQGDNSTIETQNGFRNNATIDQWNAKNDQLVTRVVT 120  
Db 61 SDARKSETTITQSGYNGADVGQGDNSTIETQNGFRNNATIDQWNAKNDQLVTRVVT 112

Qy 121 HEMAHANQASDSSVMVRQVGFNNATANQY 151

Db 113 HEMAHAGNNAALVNQASDSSVMVRQVGFNNATANQY 151

RESULT 4

US-09-543-407-5

Sequence 5, Application US/09543407

GENERAL INFORMATION:

APPLICANT: White, Aaron P.

APPLICANT: Doran, James L.

APPLICANT: Collinson, S. Karen

APPLICANT: Kay, William W.

TITLE OF INVENTION: BACTERIAL FIMBRIAL SYSTEM FOR

TITLE OF INVENTION: PRESENTATION OF HETEROLOGOUS PEPTIDE SEQUENCES

FILE REFERENCE: 920043.406

CURRENT APPLICATION NUMBER: US/09/543,407

CURRENT FILING DATE: 2000-04-05

NUMBER OF SEQ ID NOS: 59

SOFTWARE: FastSeq for Windows Version 4.0

SEQ ID NO 5

LENGTH: 151

TYPE: PRT

ORGANISM: Salmonella enteritidis

US-09-543-407-5

Query Match 89.0%; Score 690; DB 19; Length 151;

Best Local Similarity 91.4%; Pred. No. 3.4e-65;

Matches 138; Conservative 2; Mismatches 11; Indels 0; Gaps 0;

Qy 1 MKLLKVAFAAIVVSGSALAGVVPQGGGNGHNGGNSGPDSTLSIYQYGSANAALALQ 60  
Db 1 MKLLKVAFAAIVVSGSALAGVVPQGGGNGHNGGNSGPDSTLSIYQYGSANAALALQ 60  
Qy 61 SDARKSETTITQSGYNGADVGQGDNSTIETQNGFRNNATIDQWNAKNDQLVTRVVT 120  
Db 61 SDARKSETTITQSGYNGADVGQGDNSTIETQNGFRNNATIDQWNAKNDQLVTRVVT 120

Qy 121 HEMAHANQASDSSVMVRQVGFNNATANQY 151

Db 121 NNAALVNQASDSSVMVRQVGFNNATANQY 151

RESULT 5

US-08-233-642A-57

Sequence 57, Application US/08233642A

GENERAL INFORMATION:

APPLICANT: Kay, William W.

APPLICANT: Collinson, S. Karen  
APPLICANT: Clouthier, Sharon C.  
APPLICANT: Doran, James L.  
TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR SALMONELLA-  
BASED VACCINES  
TITLE OF INVENTION: BACTERIAL FIMBRIAL SYSTEM FOR  
PRESENTATION OF HETEROLOGOUS PEPTIDE SEQUENCES  
NUMBER OF SEQUENCES: 58  
CORRESPONDENCE ADDRESS:  
ADDRESS: Seed and Berry  
STREET: 6300 Columbia Center, 701 Fifth Avenue  
CITY: Seattle  
STATE: Washington  
COUNTRY: U.S.A.  
ZIP: 98104-7052  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent In Release #1.0, Version #1.25  
CURRENT APPLICATION DATA: US/06/233,642A  
FILING DATE: 26-APR-1994  
CLASSIFICATION: 424  
ATTORNEY/AGENT INFORMATION:  
NAME: King, Joshua  
REGISTRATION NUMBER: 35,570  
REFERENCE/DOCKET NUMBER: 920043.403C3  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (206) 622-4900  
TELEFAX: (206) 682-6031  
TELEX: 3723836 SEDANBERRY  
INFORMATION FOR SEQ ID NO: 57:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 151 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-233-642A-57

Query Match 88.9%; Score 689; DB 6; Length 151;  
Best Local Similarity 91.4%; Pred. No. 4.4e-65;  
Matches 138; Conservative 2; Mismatches 11; Indels 0; Gaps 0;  
QY 1 MKLLKVAAPAAIIVVSGSALAGVVPQWGGGNGHNGGNSGPDSTLSIYQYGSANAALALQ 60  
DB 1 MKLLKVAAPAAIIVVSGSALAGVVPQWGGGNGHNGGNSGPDSTLSIYQYGSANAALALQ 60  
QY 61 SDARKSETTITQSGYNGADVGCGADNSTIETQNGFRNNATIDQWNAKNDITVQYGG 120  
DB 61 SDARKSETTITQSGYNGADVGCGADNSTIETQNGFRNNATIDQWNAKNDITVQYGG 120  
QY 121 HEMAHANQATSDSSVMVROVGFNNATANQY 151  
DB 121 NNPAIVNQATSDSSVMVROVGFNNATANQY 151

RESULT 6  
US-09-543-407-12  
Sequence 12, Application US/09543407  
GENERAL INFORMATION:  
APPLICANT: White, Aaron P.  
APPLICANT: Doran, James L.  
APPLICANT: Collinson, S. Karen  
APPLICANT: Kay, William W.  
TITLE OF INVENTION: BACTERIAL FIMBRIAL SYSTEM FOR  
PRESENTATION OF HETEROLOGOUS PEPTIDE SEQUENCES  
FILE REFERENCE: 920043.406  
CURRENT APPLICATION NUMBER: US/09/543,407  
CURRENT FILING DATE: 2000-04-05  
NUMBER OF SEQ ID NOS: 59  
SOFTWARE: FastSeq for Windows Version 4.0  
SEQ ID NO 12  
LENGTH: 151

TYPE: PRT  
ORGANISM: Artificial Sequence  
FEATURE:  
OTHER INFORMATION: Recombinant Salmonella enteritidis 3b afga  
OTHER INFORMATION: sequence containing the replacement fragment  
OTHER INFORMATION: encoding PT3 from GP63 of Leishmania major.  
US-09-543-407-12  
Query Match 84.5%; Score 655; DB 19; Length 151;  
Best Local Similarity 79.8%; Pred. No. 1.9e-61;  
Matches 134; Conservative 0; Mismatches 0; Indels 34; Gaps 2;  
QY 1 MKLLKVAAPAAIIVVSGSALAGVVPQWGGGNGHNGGNSGPDSTLSIYQYGSANAALALQ 60  
DB 1 MKLLKVAAPAAIIVVSGSALAGVVPQWGGGNGHNGGNSGPDSTLSIYQYGSANAALALQ 60  
QY 61 SDARKSETTITQSGYNGADVGCGADNSTIETQNGFRNNATIDQWNAKNDITVQYGG 120  
DB 61 SDARKSETTITQSGYNGADVGCGADNSTIETQNGFRNNATIDQWNAKNDITVQYGG 120  
QY 110 -----NYDQLVTRVVTHEMAHANQATSDSSVMVROVGFNNATANQY 151  
DB 121 NNAALVNYDQLVTRVVTHEMAHA-----NNATANQY 151  
RESULT 7  
US-09-543-407-22  
Sequence 22, Application US/09543407  
GENERAL INFORMATION:  
APPLICANT: White, Aaron P.  
APPLICANT: Doran, James L.  
APPLICANT: Collinson, S. Karen  
APPLICANT: Kay, William W.  
TITLE OF INVENTION: BACTERIAL FIMBRIAL SYSTEM FOR  
PRESENTATION OF HETEROLOGOUS PEPTIDE SEQUENCES  
FILE REFERENCE: 920043.406  
CURRENT APPLICATION NUMBER: US/09/543,407  
CURRENT FILING DATE: 2000-04-05  
NUMBER OF SEQ ID NOS: 59  
SOFTWARE: FastSeq for Windows Version 4.0  
SEQ ID NO 22  
LENGTH: 151  
TYPE: PRT  
ORGANISM: Artificial Sequence  
FEATURE:  
OTHER INFORMATION: Recombinant Salmonella enteritidis 3b afga  
OTHER INFORMATION: sequence containing the replacement fragment  
OTHER INFORMATION: encoding PT3 from GP63 of Leishmania major.  
US-09-543-407-22

Query Match 78.6%; Score 609; DB 19; Length 151;  
Best Local Similarity 73.6%; Pred. No. 1.6e-56;  
Matches 128; Conservative 0; Mismatches 0; Indels 46; Gaps 2;  
QY 1 MKLLKVAAPAAIIVVSGSALAGVVPQWGGGNGHNGGNSGPDSTLSIYQYGSANAALALQ 60  
DB 1 MKLLKVAAPAAIIVVSGSALAGVVPQWGGGNGHNGGNSGPDSTLSIYQYGSANAALALQ 60  
QY 61 SDARKSETTITQSGYNGADVGCGADNSTIETQNGFRNNATIDQWNAKNDITVQYGG 120  
DB 61 SDARKSETTITQSGYNGADVGCGAD-----NYDQLVTRVVT 97  
QY 121 HEMAH-----NCTASDSSVMVROVGFNNATANQY 151  
DB 98 HEMAHADQWNAKNSDITVQYGGNNAALVNQATSDSSVMVROVGFNNATANQY 151

RESULT 8  
US-09-543-407-26  
Sequence 26, Application US/09543407  
GENERAL INFORMATION:  
APPLICANT: White, Aaron P.  
APPLICANT: Doran, James L.

APPLICANT: Collinson, S. Karen  
APPLICANT: Kay, William W.  
TITLE OF INVENTION: BACTERIAL FIMBRIAL SYSTEM FOR  
FILE REFERENCE: 920043.406  
CURRENT APPLICATION NUMBER: US/09/543,407  
CURRENT FILING DATE: 2000-04-05  
NUMBER OF SEQ ID NOS: 59  
SOFTWARE: FastSeq for Windows Version 4.0  
SEQ ID NO 26  
LENGTH: 151  
TYPE: PRT  
ORGANISM: Artificial Sequence  
FEATURE:  
OTHER INFORMATION: Recombinant Salmonella enteritidis 3b afga  
OTHER INFORMATION: sequence containing the replacement fragment  
OTHER INFORMATION: encoding PT3 from GP63 of Leishmania major.  
US-09-543-407-26

Query Match 78.6%; Score 609; DB 19; Length 151;  
Best Local Similarity 82.1%; Pred. No. 1.6e-56;  
Matches 124; Conservative 4; Mismatches 23; Indels 0; Gaps 0;

Qy 1 MLLKVAAPAAIVVSGSALAGVVPQGGGNGGNGSGPDSITLIYQYGSANAALQ 60

Db 1 MLLKVAAPAAIVVSGSALAGVVPQGGGNGGNGSGPDSITLIYQYGSANAALQ 60

Qy 61 SDARKSETTITQSGYNGADVGQADNSTIETQNGFRNATIDQWNAKNSDITVGYGG 120

Db 61 LVTRVVTHEMAHAGYNGADVGQADNSTIETQNGFRNATIDQWNAKNSDITVGYGG 120

Qy 121 HEMAHANQTASDSSVMVROVGFNNATANY 151

Db 121 NNAALVNTASDSSVMVROVGFNNATANY 151

## RESULT 9

US-09-543-407-18  
Sequence 18, Application US/09543407  
GENERAL INFORMATION:  
APPLICANT: White, Aaron P.  
APPLICANT: Doran, James L.  
APPLICANT: Collinson, S. Karen  
APPLICANT: Kay, William W.  
TITLE OF INVENTION: BACTERIAL FIMBRIAL SYSTEM FOR  
FILE REFERENCE: 920043.406  
CURRENT APPLICATION NUMBER: US/09/543,407  
CURRENT FILING DATE: 2000-04-05  
NUMBER OF SEQ ID NOS: 59  
SOFTWARE: FastSeq for Windows Version 4.0  
SEQ ID NO 18  
LENGTH: 151  
TYPE: PRT  
ORGANISM: Artificial Sequence  
FEATURE:  
OTHER INFORMATION: Recombinant Salmonella enteritidis 3b afga  
OTHER INFORMATION: sequence containing the replacement fragment  
OTHER INFORMATION: encoding PT3 from GP63 of Leishmania major.  
US-09-543-407-18

Query Match 78.5%; Score 608; DB 19; Length 151;  
Best Local Similarity 82.1%; Pred. No. 2.1e-56;  
Matches 124; Conservative 5; Mismatches 22; Indels 0; Gaps 0;

Qy 1 MLLKVAAPAAIVVSGSALAGVVPQGGGNGGNGSGPDSITLIYQYGSANAALQ 60

Db 1 MLLKVAAPAAIVVSGSALAGVVPQGGGNGGNGSGPDSITLIYQYGSANAALQ 60

Qy 61 SDARKSETTITQSGYNGADVGQADNSTIETQNGFRNATIDQWNAKNSDITVGYGG 120

Db 61 SDARKSETTITQSGYNGADVGQADNSTIETQNGFRNATIDQWNAKNSDITVGYGG 120

Qy 121 HEMAHANQTASDSSVMVROVGFNNATANY 151  
Db 121 NNAALVNTASDSSVMVROVGFNNATANY 151

## RESULT 10

US-09-543-407-20  
Sequence 20, Application US/09543407  
GENERAL INFORMATION:  
APPLICANT: White, Aaron P.  
APPLICANT: Doran, James L.  
APPLICANT: Collinson, S. Karen  
APPLICANT: Kay, William W.  
TITLE OF INVENTION: BACTERIAL FIMBRIAL SYSTEM FOR  
FILE REFERENCE: 920043.406  
CURRENT APPLICATION NUMBER: US/09/543,407  
CURRENT FILING DATE: 2000-04-05  
NUMBER OF SEQ ID NOS: 59  
SOFTWARE: FastSeq for Windows Version 4.0  
SEQ ID NO 20  
LENGTH: 151  
TYPE: PRT  
ORGANISM: Artificial Sequence  
FEATURE:  
OTHER INFORMATION: Recombinant Salmonella enteritidis 3b afga  
OTHER INFORMATION: sequence containing the replacement fragment  
OTHER INFORMATION: encoding PT3 from GP63 of Leishmania major.  
US-09-543-407-20

Query Match 78.2%; Score 606; DB 19; Length 151;  
Best Local Similarity 82.1%; Pred. No. 3.4e-56;  
Matches 124; Conservative 5; Mismatches 22; Indels 0; Gaps 0;

Qy 1 MLLKVAAPAAIVVSGSALAGVVPQGGGNGGNGSGPDSITLIYQYGSANAALQ 60

Db 1 MLLKVAAPAAIVVSGSALAGVVPQGGGNGGNGSGPDSITLIYQYGSANAALQ 60

Qy 61 SDARKSETTITQSGYNGADVGQADNSTIETQNGFRNATIDQWNAKNSDITVGYGG 120

Db 61 SDARKYDQLVTRVVTHEMAHAGGAGNSTIETQNGFRNATIDQWNAKNSDITVGYGG 120

Qy 121 HEMAHANQTASDSSVMVROVGFNNATANY 151

Db 121 NNAALVNTASDSSVMVROVGFNNATANY 151

## RESULT 11

US-09-543-407-31  
Sequence 31, Application US/09543407  
GENERAL INFORMATION:  
APPLICANT: White, Aaron P.  
APPLICANT: Doran, James L.  
APPLICANT: Collinson, S. Karen  
APPLICANT: Kay, William W.  
TITLE OF INVENTION: BACTERIAL FIMBRIAL SYSTEM FOR  
FILE REFERENCE: 920043.406  
CURRENT APPLICATION NUMBER: US/09/543,407  
CURRENT FILING DATE: 2000-04-05  
NUMBER OF SEQ ID NOS: 59  
SOFTWARE: FastSeq for Windows Version 4.0  
SEQ ID NO 31  
LENGTH: 131  
TYPE: PRT  
ORGANISM: Salmonella enteritidis  
US-09-543-407-31

Query Match 77.8%; Score 603; DB 19; Length 131;  
Best Local Similarity 90.1%; Pred. No. 5.9e-56;  
Matches 118; Conservative 2; Mismatches 11; Indels 0; Gaps 0;

Qy 21 GVPQGGGNGGNGSGPDSITLIYQYGSANAALQSDARKSETTITQSGYNGAD 80

```
Db 1 GVVPQWGGGNGGSSGPDSTLSIYQYGSANAALALQSDARKSETTITQSGYNGAD 60
QY 81 VCGGADNSTIETQNGFRNATIDOWNAKNYDQLVTRVVTHEMAHANCTASDSSVMVRQV 140
Db 61 VCGGADNSTIETQNGFRNATIDOWNAKNSDITVGGGNNALVNQTSASSSSVMVRQV 120
QY 141 GFGNNATANQY 151
Db 121 GFGNNATANQY 131

RESULT 12
US-09-543-407-28
; Sequence 28, Application US/09543407
; GENERAL INFORMATION:
; APPLICANT: White, Aaron P.
; APPLICANT: Doran, James L.
; APPLICANT: Collinson, S. Karen
; APPLICANT: Kay, William W.
; TITLE OF INVENTION: BACTERIAL FIMBRIAL SYSTEM FOR
; TITLE OF INVENTION: PRESENTATION OF HETEROLOGOUS PEPTIDE SEQUENCES
; FILE REFERENCE: 920043.406
; CURRENT APPLICATION NUMBER: US/09/543,407
; CURRENT FILING DATE: 2000-04-05
; NUMBER OF SEQ ID NOS: 59
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 28
; LENGTH: 151
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Recombinant Salmonella enteritidis 3b afga
; OTHER INFORMATION: sequence containing the replacement fragment
; OTHER INFORMATION: encoding PT3 from GP63 of Leishmania major.
US-09-543-407-28
Query Match 77.4%; Score 600; DB 19; Length 151;
Best Local Similarity 81.5%; Pred. No. 1.5e-55; Indels 0; Gaps 0;
Matches 123; Conservative 3; Mismatches 25;
; APPLICANT: White, Aaron P.
; APPLICANT: Doran, James L.
; APPLICANT: Collinson, S. Karen
; APPLICANT: Kay, William W.
; TITLE OF INVENTION: BACTERIAL FIMBRIAL SYSTEM FOR
; TITLE OF INVENTION: PRESENTATION OF HETEROLOGOUS PEPTIDE SEQUENCES
; FILE REFERENCE: 920043.406
; CURRENT APPLICATION NUMBER: US/09/543,407
; CURRENT FILING DATE: 2000-04-05
; NUMBER OF SEQ ID NOS: 59
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 28
; LENGTH: 151
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Recombinant Salmonella enteritidis 3b afga
; OTHER INFORMATION: sequence containing the replacement fragment
; OTHER INFORMATION: encoding PT3 from GP63 of Leishmania major.
US-09-543-407-28
QY 1 MLLKVAFAAIVVSGSALAGVVPQWGGGNGGNGSSGPDSTLSIYQYGSANAALALQ 60
Db 1 MLLKVAFAAIVVSGSALAGVVPQWGGGNGGNGSSGPDSTLSIYQYGSANAALALQ 60
QY 61 SDARKSETTITQSGYNGADVGGGADNSTIETQNGFRNATIDOWNAKNYDQLVTRVT 120
Db 61 SDARKSETTITQSGYNGADYDQLVTRVVTHEMAHANCTASDSSVMVRQV 151
QY 121 HEMAHANCTASDSSVMVRQVGFNNATANQY 151
Db 121 NNALVNQTSASSSSVMVRQVGFNNATANQY 151

RESULT 13
US-09-543-407-16
; Sequence 16, Application US/09543407
; GENERAL INFORMATION:
; APPLICANT: White, Aaron P.
; APPLICANT: Doran, James L.
; APPLICANT: Collinson, S. Karen
; APPLICANT: Kay, William W.
; TITLE OF INVENTION: BACTERIAL FIMBRIAL SYSTEM FOR
; TITLE OF INVENTION: PRESENTATION OF HETEROLOGOUS PEPTIDE SEQUENCES
; FILE REFERENCE: 920043.406
; CURRENT APPLICATION NUMBER: US/09/543,407
; CURRENT FILING DATE: 2000-04-05
; NUMBER OF SEQ ID NOS: 59
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 16
; LENGTH: 151
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Recombinant Salmonella enteritidis 3b afga
; OTHER INFORMATION: sequence containing the replacement fragment
; OTHER INFORMATION: encoding PT3 from GP63 of Leishmania major.
US-09-543-407-16
Query Match 77.4%; Score 600; DB 19; Length 151;
Best Local Similarity 81.5%; Pred. No. 1.5e-55; Indels 0; Gaps 0;
Matches 123; Conservative 3; Mismatches 25;
; APPLICANT: White, Aaron P.
; APPLICANT: Doran, James L.
; APPLICANT: Collinson, S. Karen
; APPLICANT: Kay, William W.
; TITLE OF INVENTION: BACTERIAL FIMBRIAL SYSTEM FOR
; TITLE OF INVENTION: PRESENTATION OF HETEROLOGOUS PEPTIDE SEQUENCES
; FILE REFERENCE: 920043.406
; CURRENT APPLICATION NUMBER: US/09/543,407
; CURRENT FILING DATE: 2000-04-05
; NUMBER OF SEQ ID NOS: 59
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 16
; LENGTH: 151
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Recombinant Salmonella enteritidis 3b afga
; OTHER INFORMATION: sequence containing the replacement fragment
; OTHER INFORMATION: encoding PT3 from GP63 of Leishmania major.
US-09-543-407-16
QY 1 MLLKVAFAAIVVSGSALAGVVPQWGGGNGGNGSSGPDSTLSIYQYGSANAALALQ 60
Db 1 MLLKVAFAAIVVSGSALAGVVPQWGGGNGGNGSSGPDSTLSIYQYGSANAALALQ 60
QY 61 SDARKSETTITQSGYNGADVGGGADNSTIETQNGFRNATIDOWNAKNYDQLVTRVT 120
Db 61 SDARKSETTITQSGYNGADYDQLVTRVVTHEMAHANCTASDSSVMVRQV 151
QY 121 HEMAHANCTASDSSVMVRQVGFNNATANQY 151
Db 121 NNALVNQTSASSSSVMVRQVGFNNATANQY 151

RESULT 14
US-09-543-407-7
; Sequence 7, Application US/09543407
; GENERAL INFORMATION:
; APPLICANT: White, Aaron P.
; APPLICANT: Doran, James L.
; APPLICANT: Collinson, S. Karen
; APPLICANT: Kay, William W.
; TITLE OF INVENTION: BACTERIAL FIMBRIAL SYSTEM FOR
; TITLE OF INVENTION: PRESENTATION OF HETEROLOGOUS PEPTIDE SEQUENCES
; FILE REFERENCE: 920043.406
; CURRENT APPLICATION NUMBER: US/09/543,407
; CURRENT FILING DATE: 2000-04-05
; NUMBER OF SEQ ID NOS: 59
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 7
; LENGTH: 151
; TYPE: PRT
; ORGANISM: Escherichia coli
; FEATURE:
; OTHER INFORMATION: Recombinant Salmonella enteritidis 3b afga
; OTHER INFORMATION: sequence containing the replacement fragment
; OTHER INFORMATION: encoding PT3 from GP63 of Leishmania major.
US-09-543-407-7
Query Match 68.8%; Score 533; DB 19; Length 151;
Best Local Similarity 70.2%; Pred. No. 2.3e-48; Indels 0; Gaps 0;
Matches 106; Conservative 18; Mismatches 27;
; APPLICANT: White, Aaron P.
; APPLICANT: Doran, James L.
; APPLICANT: Collinson, S. Karen
; APPLICANT: Kay, William W.
; TITLE OF INVENTION: BACTERIAL FIMBRIAL SYSTEM FOR
; TITLE OF INVENTION: PRESENTATION OF HETEROLOGOUS PEPTIDE SEQUENCES
; FILE REFERENCE: 920043.406
; CURRENT APPLICATION NUMBER: US/09/543,407
; CURRENT FILING DATE: 2000-04-05
; NUMBER OF SEQ ID NOS: 59
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 7
; LENGTH: 151
; TYPE: PRT
; ORGANISM: Escherichia coli
; FEATURE:
; OTHER INFORMATION: Recombinant Salmonella enteritidis 3b afga
; OTHER INFORMATION: sequence containing the replacement fragment
; OTHER INFORMATION: encoding PT3 from GP63 of Leishmania major.
US-09-543-407-7
QY 1 MLLKVAFAAIVVSGSALAGVVPQWGGGNGGNGSSGPDSTLSIYQYGSANAALALQ 60
Db 1 MLLKVAFAAIVVSGSALAGVVPQWGGGNGGNGSSGPDSTLSIYQYGSANAALALQ 60
QY 61 SDARKSETTITQSGYNGADVGGGADNSTIETQNGFRNATIDOWNAKNYDQLVTRVT 120
Db 61 TDARNSDLTITQGGGNGADVGGGSDSSIDLTORFGNSATLQWNGKSEMTVKQFGG 120
QY 121 HEMAHANCTASDSSVMVRQVGFNNATANQY 151
Db 121 GNGRAVDOTASNSSVMVRQVGFNNATANQY 151

RESULT 15
US-08-978-978-4
; Sequence 4, Application US/08978878
; GENERAL INFORMATION:
; APPLICANT: Olsen, Arne
; APPLICANT: NORMARK, Staffan
; TITLE OF INVENTION: FIBRONECTIN BINDING PROTEIN AS WELL AS ITS PREPARATION
; FILE REFERENCE: 012889-081
; CURRENT APPLICATION NUMBER: US/08/978,978
; CURRENT FILING DATE: 1997-11-26
; EARLIER APPLICATION NUMBER: SE 8801723-1
; EARLIER FILING DATE: 1988-05-06
```

```
; FEATURE:
; OTHER INFORMATION: Recombinant Salmonella enteritidis 3b afga
; OTHER INFORMATION: sequence containing the replacement fragment
; OTHER INFORMATION: encoding PT3 from GP63 of Leishmania major.
US-09-543-407-16
Query Match 74.2%; Score 575; DB 19; Length 151;
Best Local Similarity 81.5%; Pred. No. 7.2e-53; Indels 0; Gaps 0;
Matches 123; Conservative 5; Mismatches 23;
; APPLICANT: White, Aaron P.
; APPLICANT: Doran, James L.
; APPLICANT: Collinson, S. Karen
; APPLICANT: Kay, William W.
; TITLE OF INVENTION: BACTERIAL FIMBRIAL SYSTEM FOR
; TITLE OF INVENTION: PRESENTATION OF HETEROLOGOUS PEPTIDE SEQUENCES
; FILE REFERENCE: 920043.406
; CURRENT APPLICATION NUMBER: US/09/543,407
; CURRENT FILING DATE: 2000-04-05
; NUMBER OF SEQ ID NOS: 59
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 7
; LENGTH: 151
; TYPE: PRT
; ORGANISM: Escherichia coli
; FEATURE:
; OTHER INFORMATION: Recombinant Salmonella enteritidis 3b afga
; OTHER INFORMATION: sequence containing the replacement fragment
; OTHER INFORMATION: encoding PT3 from GP63 of Leishmania major.
US-09-543-407-16
QY 1 MLLKVAFAAIVVSGSALAGVVPQWGGGNGGNGSSGPDSTLSIYQYGSANAALALQ 60
Db 1 MLLKVAFAAIVVSGSALAGVVPQWGGGNGGNGSSGPDSTLSIYQYGSANAALALQ 60
QY 61 SDARKSETTITQSGYNGADVGGGADNSTIETQNGFRNATIDOWNAKNYDQLVTRVT 120
Db 61 SDARKSETTITQSGYNGADVGGGADNSTIETQNGFRNATIDOWNAKNSDITVGGYGG 120
QY 121 HEMAHANCTASDSSVMVRQVGFNNATANQY 151
Db 121 NNALVNQTSASSSSVMVRQVGFNNATANQY 151

RESULT 14
US-09-543-407-7
; Sequence 7, Application US/09543407
; GENERAL INFORMATION:
; APPLICANT: White, Aaron P.
; APPLICANT: Doran, James L.
; APPLICANT: Collinson, S. Karen
; APPLICANT: Kay, William W.
; TITLE OF INVENTION: BACTERIAL FIMBRIAL SYSTEM FOR
; TITLE OF INVENTION: PRESENTATION OF HETEROLOGOUS PEPTIDE SEQUENCES
; FILE REFERENCE: 920043.406
; CURRENT APPLICATION NUMBER: US/09/543,407
; CURRENT FILING DATE: 2000-04-05
; NUMBER OF SEQ ID NOS: 59
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 7
; LENGTH: 151
; TYPE: PRT
; ORGANISM: Escherichia coli
; FEATURE:
; OTHER INFORMATION: Recombinant Salmonella enteritidis 3b afga
; OTHER INFORMATION: sequence containing the replacement fragment
; OTHER INFORMATION: encoding PT3 from GP63 of Leishmania major.
US-09-543-407-7
Query Match 68.8%; Score 533; DB 19; Length 151;
Best Local Similarity 70.2%; Pred. No. 2.3e-48; Indels 0; Gaps 0;
Matches 106; Conservative 18; Mismatches 27;
; APPLICANT: White, Aaron P.
; APPLICANT: Doran, James L.
; APPLICANT: Collinson, S. Karen
; APPLICANT: Kay, William W.
; TITLE OF INVENTION: BACTERIAL FIMBRIAL SYSTEM FOR
; TITLE OF INVENTION: PRESENTATION OF HETEROLOGOUS PEPTIDE SEQUENCES
; FILE REFERENCE: 920043.406
; CURRENT APPLICATION NUMBER: US/09/543,407
; CURRENT FILING DATE: 2000-04-05
; NUMBER OF SEQ ID NOS: 59
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 7
; LENGTH: 151
; TYPE: PRT
; ORGANISM: Escherichia coli
; FEATURE:
; OTHER INFORMATION: Recombinant Salmonella enteritidis 3b afga
; OTHER INFORMATION: sequence containing the replacement fragment
; OTHER INFORMATION: encoding PT3 from GP63 of Leishmania major.
US-09-543-407-7
QY 1 MLLKVAFAAIVVSGSALAGVVPQWGGGNGGNGSSGPDSTLSIYQYGSANAALALQ 60
Db 1 MLLKVAFAAIVVSGSALAGVVPQWGGGNGGNGSSGPDSTLSIYQYGSANAALALQ 60
QY 61 SDARKSETTITQSGYNGADVGGGADNSTIETQNGFRNATIDOWNAKNYDQLVTRVT 120
Db 61 TDARNSDLTITQGGGNGADVGGGSDSSIDLTORFGNSATLQWNGKSEMTVKQFGG 120
QY 121 HEMAHANCTASDSSVMVRQVGFNNATANQY 151
Db 121 GNGRAVDOTASNSSVMVRQVGFNNATANQY 151

RESULT 15
US-08-978-978-4
; Sequence 4, Application US/08978878
; GENERAL INFORMATION:
; APPLICANT: Olsen, Arne
; APPLICANT: NORMARK, Staffan
; TITLE OF INVENTION: FIBRONECTIN BINDING PROTEIN AS WELL AS ITS PREPARATION
; FILE REFERENCE: 012889-081
; CURRENT APPLICATION NUMBER: US/08/978,978
; CURRENT FILING DATE: 1997-11-26
; EARLIER APPLICATION NUMBER: SE 8801723-1
; EARLIER FILING DATE: 1988-05-06
```

```
; EARLIER APPLICATION NUMBER: US 07/347,189
; EARLIER FILING DATE: 1989-05-04
; EARLIER APPLICATION NUMBER: US 07/789,437
; EARLIER FILING DATE: 1991-11-06
; EARLIER APPLICATION NUMBER: US 07/970,846
; EARLIER FILING DATE: 1992-11-03
; EARLIER APPLICATION NUMBER: US 08/187,865
; EARLIER FILING DATE: 1994-01-28
; EARLIER APPLICATION NUMBER: US 08/318,519
; EARLIER FILING DATE: 1994-10-05
; EARLIER APPLICATION NUMBER: US 08/495,959
; EARLIER FILING DATE: 1995-06-28
; NUMBER OF SEQ ID NOS: 10
; SOFTWARE: Patentin Ver. 2.0
; SEQ ID NO 4
; LENGTH: 151
; TYPE: PRT
; ORGANISM: Escherichia coli
US-08-978-878-4

Query Match      68.4%; Score 530; DB 13; Length 151;
Best Local Similarity 69.5%; Pred. No. 4.8e-48;
Matches 105; Conservative 19; Mismatches 27; Indels 0; Gaps 0;

QY      1 MLLKVAFAAIVVSGSALAGVVPQGGGNGHGGNSGDPSTLSIYCYGSANAALAQ 60
Db      1 MLLKVAFAAIVFSGSAGVAVVPQGGGNGHGGNSGDPSTLSIYCYGSANAALAQ 60
QY      61 SPARKSETTITQSGYNGADVGQGGADNSTIELTQNGFRNNATIDQNNKDYDQVTRVYT 120
Db      61 TDARNSDLTITQGGGNGADVGQGGSDSSIDLITQRFNGNSATLDQNGKNSMTVKQFGG 120
QY      121 HEMAHANQTASDSSVMVROVFGNNATANQY 151
Db      121 GNGAAVDQTASNSVNVTCVFGNNATAHQY 151
```

Search completed: August 2, 2004, 15:26:45  
Job time : 167.9 secs



GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: August 2, 2004, 14:49:38 ; Search time 17.8 Seconds  
(without alignments)  
888.146 Million cell updates/sec

Title: US-09-543-407-24  
Perfect score: 775  
Sequence: 1 MKLLKVAFAAIVVSGSALA.....DSSVMVRQVGFNNATANQY 151

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 601315 seqs, 104695340 residues

Total number of hits satisfying chosen parameters: 601315

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Pending Patents AA New.\*  
1: /cgn2\_6/prodata/2/paa/PCT\_NEW\_COMB.pep.\*  
2: /cgn2\_6/prodata/2/paa/US05\_NEW\_COMB.pep.\*  
3: /cgn2\_6/prodata/2/paa/US07\_NEW\_COMB.pep.\*  
4: /cgn2\_6/prodata/2/paa/US08\_NEW\_COMB.pep.\*  
5: /cgn2\_6/prodata/2/paa/US09\_NEW\_COMB.pep.\*  
6: /cgn2\_6/prodata/2/paa/US10\_NEW\_COMB.pep.\*  
7: /cgn2\_6/prodata/2/paa/US60\_NEW\_COMB.pep.\*

pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	530	68.4	151	5	US-09-741-873C-4
2	452	58.3	131	5	US-09-741-873C-2
3	91	11.7	719	5	US-09-248-796A-17559
4	90.5	11.7	258	6	US-10-425-115-300390
5	90.5	11.7	295	6	US-10-425-115-312468
6	90.5	11.7	299	6	US-10-170-205B-35751
7	90.5	11.7	520	6	US-10-479-638-21
8	90.5	11.7	1203	6	US-10-170-205E-741
9	90.5	11.7	1327	1	PCT-US04-02338-49
10	89.5	11.5	256	6	US-10-425-115-301334
11	85.5	11.0	388	5	US-09-248-796A-17306
12	85	11.0	535	1	PCT-US04-21492-88
13	84	10.8	376	6	US-10-491-733-2
14	83.5	10.8	179	6	US-10-425-115-346132
15	83.5	10.8	573	7	US-60-565-632-7907
16	83.5	10.8	573	7	US-60-579-062-7907
17	83	10.7	889	5	US-09-952-267B-15
18	83	10.7	889	6	US-10-872-768-15
19	83	10.7	889	6	US-10-872-769-15
20	81.5	10.5	508	6	US-10-425-115-285216
21	81.5	10.5	532	6	US-10-425-115-285214
22	81	10.5	131	7	US-60-565-632-11109
23	81	10.5	131	7	US-60-579-062-11109
24	81	10.5	201	6	US-10-425-115-309662
25	81	10.5	376	6	US-10-425-115-342526
26	81	10.5	382	6	US-10-771-241-299

27 81 10.5 659 1 PCT-US04-09385-4 Sequence 4, Appli  
28 81 10.5 659 6 US-10-809-790-4 Sequence 4, Appli  
29 80.5 10.4 246 6 US-10-854-439-511 Sequence 511, App  
30 80.5 10.4 389 6 US-10-724-972A-4638 Sequence 4638, Ap  
31 80.5 10.4 891 7 US-60-566-425-574 Sequence 574, App  
32 80.5 10.4 891 7 US-60-576-812-606 Sequence 606, App  
33 80.5 10.4 894 7 US-60-566-425-570 Sequence 570, App  
34 80.5 10.4 894 7 US-60-566-425-571 Sequence 571, App  
35 80.5 10.4 894 7 US-60-566-425-575 Sequence 575, App  
36 80.5 10.4 894 7 US-60-566-425-578 Sequence 578, App  
37 80.5 10.4 894 7 US-60-576-812-602 Sequence 602, App  
38 80.5 10.4 894 7 US-60-576-812-603 Sequence 603, App  
39 80.5 10.4 894 7 US-60-576-812-607 Sequence 607, App  
40 80.5 10.4 894 7 US-60-576-812-610 Sequence 610, App  
41 80.5 10.4 898 6 US-10-170-205E-12329 Sequence 12329, A  
42 80.5 10.4 898 7 US-60-566-425-577 Sequence 577, App  
43 80.5 10.4 898 7 US-60-576-812-609 Sequence 609, App  
44 79 10.2 132 6 US-10-425-115-351875 Sequence 351875,  
45 79 10.2 218 5 US-09-248-796A-14423 Sequence 14423, A

#### ALIGNMENTS

RESULT 1  
US-09-741-873C-4  
; Sequence 4, Application US/09741873C  
; GENERAL INFORMATION:  
; APPLICANT: Normark, Staffan  
; TITLE OF INVENTION: Fibronectin Binding Protein As Well As Its Preparation  
; FILE REFERENCE: 012889-084  
; CURRENT APPLICATION NUMBER: US/09/741.873C  
; CURRENT FILING DATE: 2000-12-22  
; PRIOR APPLICATION NUMBER: SE 8801723-1  
; PRIOR FILING DATE: 1998-05-06  
; PRIOR APPLICATION NUMBER: US 08/978,878  
; PRIOR FILING DATE: 1997-11-26  
; PRIOR APPLICATION NUMBER: US 07/347,189  
; PRIOR FILING DATE: 1989-05-04  
; PRIOR APPLICATION NUMBER: US 07/789,437  
; PRIOR FILING DATE: 1991-11-06  
; PRIOR APPLICATION NUMBER: US 07/970,846  
; PRIOR FILING DATE: 1992-11-03  
; PRIOR APPLICATION NUMBER: US 08/187,865  
; PRIOR FILING DATE: 1994-01-28  
; PRIOR APPLICATION NUMBER: US 08/318,519  
; PRIOR FILING DATE: 1994-10-05  
; NUMBER OF SEQ ID NOS: 11  
; SOFTWARE: Patent in version 3.0  
; SEQ ID NO 4  
; LENGTH: 151  
; TYPE: PRT  
; ORGANISM: Escherichia coli  
US-09-741-873C-4

Query Match 68.4%; Score 530; DB 5; Length 151;  
Best Local Similarity 69.5%; Pred. No. 1.8e-39;  
Matches 105; Conservative 19; Mismatches 27; Indels 0; Gaps 0;

QY 1 MKLLKVAFAAIVVSGSALAAGVFPQGGGNGHNGSSGPDSTLSIYQGSANAALAQ 60  
DB 1 MKLLKVAFAAIVFSGSAVAGVFPQYGGGNGHNGSSGNSLNLYIYGGNSALALQ 60  
QY 61 SDARKSETTITQSGYNGADYVQGGADNSTIELTQNGFRNNATIDQWNAKNDQLVTRVVT 120  
DB 61 TDARNSDLTITQHGNGGADYVQGGSDSSIDLTKQFGNSATLQWNGKNSMTVKQFGG 120  
QY 121 HEMAHANTASDSSVMVRQVGFNNATANQY 151  
DB 121 GNGAAVDQTSNNSVNVTVQVGFNNATAHQY 151

RESULT 2  
US-09-741-873C-2  
; Sequence 2, Application US/09741873C  
; GENERAL INFORMATION:  
; APPLICANT: Normark, Staffan  
; APPLICANT: Olsen, Arne  
; TITLE OF INVENTION: Fibronectin Binding Protein As Well As Its Preparation  
; FILE REFERENCE: 012889-084  
; CURRENT APPLICATION NUMBER: US/09/741,873C  
; CURRENT FILING DATE: 2000-12-22  
; PRIOR APPLICATION NUMBER: SE 8801723-1  
; PRIOR FILING DATE: 1998-05-06  
; PRIOR APPLICATION NUMBER: US 08/978,878  
; PRIOR FILING DATE: 1997-11-26  
; PRIOR APPLICATION NUMBER: US 07/347,189  
; PRIOR FILING DATE: 1989-05-04  
; PRIOR APPLICATION NUMBER: US 07/789,437  
; PRIOR FILING DATE: 1991-11-06  
; PRIOR APPLICATION NUMBER: US 07/970,846  
; PRIOR FILING DATE: 1992-11-03  
; PRIOR APPLICATION NUMBER: US 08/187,865  
; PRIOR FILING DATE: 1994-01-28  
; PRIOR APPLICATION NUMBER: US 08/318,519  
; PRIOR FILING DATE: 1994-10-05  
; NUMBER OF SEQ ID NOS: 11  
; SOFTWARE: Patent in version 3.0  
; SEQ ID NO 2  
; LENGTH: 131  
; TYPE: PRT  
; ORGANISM: Escherichia coli  
US-09-741-873C-2

Query Match 58.3%; Score 452; DB 5; Length 131;  
Best Local Similarity 66.4%; Pred. No. 1.2e-32;  
Matches 87; Conservative 18; Mismatches 26; Indels 0; Gaps 0;  
  
QY 21 GVVPQGGGNNHGGSSGPDSTLSIYQGSANAALALQSDARKSETTITQSGYGNAGD 80  
DB 1 GVVPQGGGNNHGGGNNGPNSELNTYQGGGNSALALQTDARNSDLTITQGGGNGAD 60  
  
QY 81 VGGGADNSTIELQNGFRNATIDQNAKNDYDQVTRVVTHEMAHANOTASDSSVMVRQV 140  
DB 61 VGGGSDSSIDLQRFNGSATLQNGKNSMTVQFGGNGAAYDQTAASNSVNVITQV 120  
  
QY 141 GFGNNATAOY 151  
DB 121 GFGNNATAHOY 131

RESULT 3  
US-09-248-796A-17559  
; Sequence 17559, Application US/09248796A  
; GENERAL INFORMATION:  
; APPLICANT: Keith Weinstock et al  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO CANDIDA ALBICAN  
; TITLE OF INVENTION: FOR DIAGNOSTICS AND THERAPEUTICS  
; FILE REFERENCE: 107196-132  
; CURRENT APPLICATION NUMBER: US/09/248,796A  
; CURRENT FILING DATE: 1999-02-12  
; PRIOR APPLICATION NUMBER: US 60/074,725  
; PRIOR FILING DATE: 1998-02-13  
; PRIOR APPLICATION NUMBER: US 60/096,409  
; PRIOR FILING DATE: 1998-08-13  
; NUMBER OF SEQ ID NOS: 28208  
; SEQ ID NO 17559  
; LENGTH: 719  
; TYPE: PRT  
; ORGANISM: Candida albicans  
US-09-248-796A-17559

Query Match 11.7%; Score 91; DB 5; Length 719;  
Best Local Similarity 23.7%; Pred. No. 6.2;  
Matches 28; Conservative 25; Mismatches 59; Indels 6; Gaps 2;

QY 30 GNHGGSSGPDSTLSIYQGSANAALALQSDARKSETTITQSGYGNAGDVGQGGADNST 89  
DB 519 GNNGSGNSGTTNNGNNYNNKSIKKNEIDGDDDLNPTISITNN--TGLTNNNSKSPA 575  
  
QY 90 IEITQNGFRNATIDQNAKNDYDQVTRVVTHEMAHANOTASDSSVMVRQVGFNNAT 147  
DB 576 KSKKGNFNNNSNS---ALNLDKSKLKINTNEITNSETTSNSSSPFVNLNHHGRSS 630  
  
RESULT 4  
US-10-425-115-300390  
; Sequence 300390, Application US/10425115  
; GENERAL INFORMATION:  
; APPLICANT: La Rosa, Thomas J.  
; APPLICANT: Kovalic, David K.  
; APPLICANT: Zhou, Yihua  
; APPLICANT: Cao, Yongwei  
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With  
; TITLE OF INVENTION: Plants  
; FILE REFERENCE: 38-21(53222)B  
; CURRENT APPLICATION NUMBER: US/10/425,115  
; CURRENT FILING DATE: 2003-04-28  
; NUMBER OF SEQ ID NOS: 369326  
; SEQ ID NO 300390  
; LENGTH: 258  
; TYPE: PRT  
; ORGANISM: Zea mays  
; FEATURE:  
; NAME/KEY: unsure  
; LOCATION: (1)..(258)  
; OTHER INFORMATION: unsure at all xaa locations  
; FEATURE:  
; OTHER INFORMATION: Clone ID: MRT4577\_37025C.1.pep  
US-10-425-115-300390

Query Match 11.7%; Score 90.5; DB 6; Length 258;  
Best Local Similarity 27.9%; Pred. No. 2.1;  
Matches 29; Conservative 15; Mismatches 39; Indels 21; Gaps 4;  
  
QY 17 SALAGVVPQGGGNNHGGSSGPDSTLSIYQGSANAALALQSDARKSETTITQSG-- 74  
DB 78 SSIAAG-----GGGGGQGGGGGTNGSGSGSGSGSGSSSTAASGSSGNYADAEKGAG 132  
  
QY 75 -----YNGCA--DVQGGADNSTIEL--TQNGFRNATID 104  
DB 133 GGMGGGANGAYGSGAGGKGEGVSGVALAPSSNGYNGGAAD 176

RESULT 5  
US-10-425-115-312468  
; Sequence 312468, Application US/10425115  
; GENERAL INFORMATION:  
; APPLICANT: La Rosa, Thomas J.  
; APPLICANT: Kovalic, David K.  
; APPLICANT: Zhou, Yihua  
; APPLICANT: Cao, Yongwei  
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With  
; TITLE OF INVENTION: Plants  
; FILE REFERENCE: 38-21(53222)B  
; CURRENT APPLICATION NUMBER: US/10/425,115  
; CURRENT FILING DATE: 2003-04-28  
; NUMBER OF SEQ ID NOS: 369326  
; SEQ ID NO 312468  
; LENGTH: 295  
; TYPE: PRT  
; ORGANISM: Zea mays  
; FEATURE:  
; OTHER INFORMATION: Clone ID: MRT4577\_48027C.1.pep  
US-10-425-115-312468

Query Match 11.7%; Score 90.5; DB 6; Length 295;  
Best Local Similarity 33.3%; Pred. No. 2.5;

Matches 29; Conservative 9; Mismatches 44; Indels 5; Gaps 2;  
 QY 26 MGGGNGHNGSGPDSTLSIYQYGSANAALQSDARKSETTITQSGYNGADVGGA 85  
 Db 138 YGGGYSGGYSYG-GYAANGYGVSGSGNYSNASGGYSGS-----DGYNGAASGGYA 192  
 QY 86 DNSTIETQNGFRNNATIDQWNAKYD 112  
 Db 193 NNLSSGYNNGRYTIGSDGNTGGYN 219

RESULT 6  
 US-10-170-205E-35751  
 ; Sequence 35751, Application US/10170205E  
 ; GENERAL INFORMATION:  
 ; APPLICANT: ADAMS, Mark  
 ; TITLE OF INVENTION: DEVICES, SUCH AS ARRAYS, COMPRISED OF HUMAN PROTEINS OR PROTEIN  
 ; TITLE OF INVENTION: CAPTURE AGENTS, AND USES THEREOF  
 ; FILE REFERENCE: CL001381  
 ; CURRENT APPLICATION NUMBER: US/10/170,205E  
 ; CURRENT FILING DATE: 2002-06-13  
 ; NUMBER OF SEQ ID NOS: 40312  
 ; SOFTWARE: PatentIn version 3.2  
 ; SEQ ID NO 35751  
 ; LENGTH: 299  
 ; TYPE: PRT  
 ; ORGANISM: Homo sapiens  
 US-10-170-205E-35751

Query Match 11.7%; Score 90.5; DB 6; Length 299;  
 Best Local Similarity 28.4%; Pred. No. 2.5;  
 Matches 42; Conservative 16; Mismatches 61; Indels 29; Gaps 6;  
 QY 6 VAAFAAI-VVSGSALAGVVPQWGGGNGHNGSGPDSTLSIYQYGSANAALQSDAR 64  
 Db 99 VAAAPVPAVSTSSAAGVAPNPAGSGNNPSSSSPTSS-SSSPSPSPGSLAESPEAA 157  
 QY 65 KSETTIT---QSGYNGADVGQADNSTIETQNGFRNNATIDQWNAKYDQLVTRVVT 120  
 Db 158 GVSSTAPLPGAGPTGTGPAVSGALRELLACRNG-----DVSrvKRLVDA--- 204  
 QY 121 HEMAHANOTASD-----SSVMVRQVGF 143  
 Db 205 -----ANVNKDMAGRKSSPLHFAAGFG 227

RESULT 7  
 US-10-479-638-21  
 ; Sequence 21, Application US/10479638  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Don A. Roth  
 ; APPLICANT: Randolph V. Lewis  
 ; TITLE OF INVENTION: The University of Wyoming  
 ; TITLE OF INVENTION: Expression of Spider Silk Proteins in Higher Plants  
 ; FILE REFERENCE: WYO.02-0004US  
 ; CURRENT APPLICATION NUMBER: US/10/479,638  
 ; CURRENT FILING DATE: 2003-12-03  
 ; PRIOR APPLICATION NUMBER: PCT/US02/18256  
 ; PRIOR FILING DATE: 2002-06-06  
 ; PRIOR APPLICATION NUMBER: 60/296,184  
 ; PRIOR FILING DATE: 2001-06-06  
 ; NUMBER OF SEQ ID NOS: 54  
 ; SOFTWARE: FastSeq for Windows Version 3.0  
 ; SEQ ID NO 21  
 ; LENGTH: 520  
 ; TYPE: PRT  
 ; ORGANISM: Argiope trifasciata  
 US-10-479-638-21

Query Match 11.7%; Score 90.5; DB 6; Length 520;  
 Best Local Similarity 24.6%; Pred. No. 4.6;  
 Matches 34; Conservative 27; Mismatches 64; Indels 13; Gaps 4;

QY 15 SGALLAGVVPQWGGGNGHNGSGPDSTLSIYQYGSANAALQSDARKSET-TITQS 73  
 Db 347 AGAGAAAAAGAGAGAGGGYGVAGGSS-----ISYGATSSSATSSSTASSRSGIVTSG 402  
 QY 74 GYNGADVGQADNSTIETQNGFRNNATIDQWNAKYDQLVTRVVTHEMAHANOTASDS 133  
 Db 403 CYGAGAAAGAGAGAGAGSYSG-----SISRLSSA---EAVNRVSSNIGAVASGGASAL 454  
 QY 134 SVMVRQVGFNNATANQY 151  
 Db 455 PGVISNIFGVSSSAGSY 472

RESULT 8  
 US-10-170-205E-741  
 ; Sequence 741, Application US/10170205E  
 ; GENERAL INFORMATION:  
 ; APPLICANT: ADAMS, Mark  
 ; TITLE OF INVENTION: DEVICES, SUCH AS ARRAYS, COMPRISED OF HUMAN PROTEINS OR PROTEIN  
 ; TITLE OF INVENTION: CAPTURE AGENTS, AND USES THEREOF  
 ; FILE REFERENCE: CL001381  
 ; CURRENT APPLICATION NUMBER: US/10/170,205E  
 ; CURRENT FILING DATE: 2002-06-13  
 ; NUMBER OF SEQ ID NOS: 40312  
 ; SOFTWARE: PatentIn version 3.2  
 ; SEQ ID NO 741  
 ; LENGTH: 1203  
 ; TYPE: PRT  
 ; ORGANISM: Homo sapiens  
 US-10-170-205E-741

Query Match 11.7%; Score 90.5; DB 6; Length 1203;  
 Best Local Similarity 28.4%; Pred. No. 13;  
 Matches 42; Conservative 16; Mismatches 61; Indels 29; Gaps 6;  
 QY 6 VAAFAAI-VVSGSALAGVVPQWGGGNGHNGSGPDSTLSIYQYGSANAALQSDAR 64  
 Db 99 VAAAPVPAVSTSSAAGVAPNPAGSGNNPSSSSPTSS-SSSPSPSPGSLAESPEAA 157  
 QY 65 KSETTIT---QSGYNGADVGQADNSTIETQNGFRNNATIDQWNAKYDQLVTRVVT 120  
 Db 158 GVSSTAPLPGAGPTGTGPAVSGALRELLACRNG-----DVSrvKRLVDA--- 204  
 QY 121 HEMAHANOTASD-----SSVMVRQVGF 143  
 Db 205 -----ANVNKDMAGRKSSPLHFAAGFG 227

RESULT 9  
 PCT-US04-02338-49  
 ; Sequence 49, Application PC/TUS0402338  
 ; GENERAL INFORMATION:  
 ; APPLICANT: EXELIXIS, INC.  
 ; TITLE OF INVENTION: MAPCAXS AS MODIFIERS OF THE APC AND AXIN PATHWAYS AND METHODS OF  
 ; TITLE OF INVENTION: USE  
 ; FILE REFERENCE: EX04-003C-PC  
 ; CURRENT APPLICATION NUMBER: PCT/US04/02338  
 ; CURRENT FILING DATE: 2004-01-28  
 ; PRIOR APPLICATION NUMBER: US60/443,484  
 ; PRIOR FILING DATE: 2003-01-29  
 ; PRIOR APPLICATION NUMBER: US60/447,358  
 ; PRIOR FILING DATE: 2003-02-11  
 ; PRIOR APPLICATION NUMBER: US60/461,789  
 ; PRIOR FILING DATE: 2003-04-10  
 ; PRIOR APPLICATION NUMBER: US60/470,684  
 ; PRIOR FILING DATE: 2003-05-14  
 ; PRIOR APPLICATION NUMBER: US60/479,650  
 ; PRIOR FILING DATE: 2003-06-19  
 ; NUMBER OF SEQ ID NOS: 54  
 ; SOFTWARE: PatentIn version 3.2  
 ; SEQ ID NO 49  
 ; LENGTH: 1327  
 ; TYPE: PRT



```

; APPLICANT: Weislo, Laura J.
; TITLE OF INVENTION: Nucleic Acid Molecules from Rice Encoding Proteins for Abiotic St
; TITLE OF INVENTION: Resistance, Yield, Disease Resistance and Nutritional Quality an
; FILE REFERENCE: 60127WOPCT
; CURRENT APPLICATION NUMBER: US/10/491,733
; PRIOR FILING DATE: 2004-04-05
; PRIOR APPLICATION NUMBER: 60/334,501
; PRIOR FILING DATE: 2001-11-30
; NUMBER OF SEQ ID NOS: 64
; SOFTWARE: Patentin version 3.1
; SEQ ID NO 2
; LENGTH: 376
; TYPE: PRT
; ORGANISM: Oryza sativa
US-10-491-733-2

```

```

Query Match      10.8%; Score 84; DB 6; Length 376;
Best Local Similarity 31.5%; Pred. No. 12;
Matches 34; Conservative 15; Mismatches 39; Indels 20; Gaps 6;

QY 17 SALAGVVPQWGGGNGHNGSSGPDSTLSIYQYGSANAALALQS--DARKSETTITQSG 74
Db 50 SALAQ-----GGGG--GGGGSGFGKD-----FGSMWDELLRISWTAEBSQMASASG 95

QY 75 YNGGADVGGGADNSTIELTQNGFRNNATIDOWNAKYDQLVTRVTHE 122
Db 96 SAAGVAVGAPPTSLQ-RQGSLLTLPRTL---SAKTVDEWRLVRDE 139

```

```

RESULT 14
US-10-425-115-346132
; Sequence 346132, Application US/10425115
; GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J.
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yinhua
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
; TITLE OF INVENTION: Plants
; FILE REFERENCE: 38-21(53222)B
; CURRENT APPLICATION NUMBER: US/10/425,115
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 369326
; SEQ ID NO 346132
; LENGTH: 179
; TYPE: PRT
; ORGANISM: Zea mays
; FEATURE:
; OTHER INFORMATION: Clone ID: MRT4577_78839C.1.pep
US-10-425-115-346132

```

```

Query Match      10.8%; Score 83.5; DB 6; Length 179;
Best Local Similarity 30.4%; Pred. No. 5.8;
Matches 28; Conservative 8; Mismatches 33; Indels 23; Gaps 4;

QY 19 LAGVFPQWGGG-----NHHGGGSSGPDSTLSIYQYGSANAALALQSDARKSETTI 70
Db 64 LDGLLSGLGGGGGLDGLLGTGGGGGKNAQADSGNAGGSGNA----QEDSGNAQ--- 116

QY 71 TQSGYNGADVGGGADNSTIELTQNGFRNNAT 102
Db 117 EESSNNQACAGAGA-----ENGAAANGT 140

```

```

RESULT 15
US-60-565-632-7907
; Sequence 7907, Application US/60565632
; GENERAL INFORMATION:
; APPLICANT: Monsanto Technology, LLC
; APPLICANT: Baum, James A.
; APPLICANT: Kovalic, David K.
; APPLICANT: Larosa, Thomas J.
; APPLICANT: Lu, Maolong

```

```

; APPLICANT: Munyikwa, Tichifa R. I.
; APPLICANT: Roberts, James K.
; APPLICANT: Wu, Wei
; APPLICANT: Zhang, Bei
; TITLE OF INVENTION: Methods for Genetic Control of Insect Infestations in Plants and
; TITLE OF INVENTION: Compositions Thereof
; FILE REFERENCE: 38-21(53403)B
; CURRENT APPLICATION NUMBER: US/60/565,632
; CURRENT FILING DATE: 2004-04-27
; NUMBER OF SEQ ID NOS: 15449
; SOFTWARE: Patentin version 3.2
; SEQ ID NO 7907
; LENGTH: 573
; TYPE: PRT
; ORGANISM: Diabrotica virgifera
US-60-565-632-7907

```

```

Query Match      10.8%; Score 83.5; DB 7; Length 573;
Best Local Similarity 27.8%; Pred. No. 22;
Matches 45; Conservative 15; Mismatches 73; Indels 29; Gaps 7;

QY 10 AAIVVSGSALAGVVPQWGGGNGHNGSSGPDSTLSIYQYGSANAALA-----LQSD 62
Db 155 AADNNGSADAA-----QGNDRNAAENANANADAQTDAAG--GSANEANANANADAQND 208

QY 63 ARKSETTITQSGYNGADVGGGADN-STIELTON-----GFRNATID-----QWNAK 109
Db 209 AAQANDNGAAAENNGNADAAQGTDEAAEENSGNENGTGAENNANADAQTDVAGGSTNEA 268

QY 110 NYDQLVTRVVTHEMAHQ---TASDSSVMVROVGFGNATA 148
Db 269 NAENNANADVQNDAAQAENGAENSGNADAAQGTDTNGAAA 310

```

```

Search completed: August 2, 2004, 15:29:54
Job time : 17.8 secs

```

This Page Blank (uspto)



Db 121 NNAALVNQTASDSSVMVRQVGFNNATANQY 151

RESULT 2

AI0635

major curlin chain precursor [imported] - Salmonella enterica subsp. enterica serovar Typhimurium

C:Species: Salmonella enterica subsp. enterica serovar Typhi

A:Note: this species has also been called Salmonella typhi

C:Date: 09-Nov-2001 #sequence\_revision 09-Nov-2001 #text\_change 18-Nov-2002

C:Accession: AI0635

R:Parkhill, J.; Dougan, G.; James, K.D.; Thomson, N.R.; Pickard, D.; Wain, J.; Churcher, T.; Connerton, P.; Cronin, A.; Davis, P.; Davies, R.M.; Dowd, L.; White, N.; Farrar, S.; Moule, S.; O'Garra, P.

A:Authors: Parkhill, J.; Dougan, G.; James, K.D.; Thomson, N.R.; Pickard, D.; Wain, J.; Churcher, T.; Connerton, P.; Cronin, A.; Davis, P.; Davies, R.M.; Dowd, L.; White, N.; Farrar, S.; Moule, S.; O'Garra, P.

A:Title: Complete genome sequence of a multiple drug resistant Salmonella enterica serovar Typhimurium

A:Reference number: AB0502; MUID:21534947; PMID:11677608

A:Accession: AI0635

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-151 <PAR>

A:Cross-references: GB:AL513382; PIDN:CAD08268.1; PID:g16502315; GSPDB:GN00176

C:Genetics:

A:Gene: STV1181

Query Match 89.0%; Score 690; DB 2; Length 151;

Best Local Similarity 91.4%; Pred. No. 4.7e-51;

Matches 138; Conservative 2; Mismatches 11; Indels 0; Gaps 0;

Qy 1 MKLLKVAAPAAIVVSGSALAGVVPQWGGGNGHNGGNSGPDSTLSIYQYGSANAALALQ 60

Db 1 MKLLKVAAPAAIVVSGSALAGVVPQWGGGNGHNGGNSGPDSTLSIYQYGSANAALALQ 60

Qy 61 SDARKSETTITQSGYNGADVGAGDNSTIELTQNGFRNNATIDQWAKNYDQLVTRVVT 120

Db 61 SDARKSETTITQSGYNGADVGAGDNSTIELTQNGFRNNATIDQWAKNYDQLVTRVVT 120

Qy 121 HEMAHANQTASDSSVMVRQVGFNNATANQY 151

Db 121 NNAALVNQTASDSSVMVRQVGFNNATANQY 151

RESULT 3

S70788

curlin protein csgA precursor - Escherichia coli (strain K-12)

N:Alternate names: csgA protein; major curlin protein

C:Species: Escherichia coli

C:Date: 12-Feb-1998 #sequence\_revision 20-Feb-1998 #text\_change 01-Mar-2002

C:Accession: S70788; G64846; S31202; S34560; S34559

R:Hammar, M.; Arngqvist, A.; Bian, Z.; Olsen, A.; Normark, S.

Mol. Microbiol. 18, 661-670, 1995

A:Title: Expression of two csg operons is required for production of fibronectin- and curli

A:Reference number: S70783; MUID:96414468; PMID:8817489

A:Accession: S70788

A:Status: nucleic acid sequence not shown; translation not shown

A:Molecule type: DNA

A:Residues: 1-151 <HAM>

A:Cross-references: EMBL:X90754; NID:g1147558; PIDN:CAA62282.1; PID:g1147564

A:Experimental source: strain K12, substrain W3110

A:Note: the nucleotide sequence was submitted to the EMBL Data Library, August 1995

R:Blattner, F.R.; Plunkett III, G.; Bloch, C.A.; Perna, N.T.; Burland, V.; Riley, M.; Cohen, D.; Rose, D.J.; Mau, B.; Shaoh, Y.

Science 277, 1453-1462, 1997

A:Title: The complete genome sequence of Escherichia coli K-12.

A:Reference number: A64720; MUID:97426617; PMID:9278503

A:Accession: G64846

A:Status: nucleic acid sequence not shown; translation not shown

A:Molecule type: DNA

A:Residues: 1-151 <BLAT>

A:Cross-references: GB:AE000205; GB:U00096; NID:g1787265; PIDN:AAC47126.1; PID:g1787279;

A:Experimental source: strain K-12, substrain MG1655

R:Olsen, A.; Arngqvist, A.; Hammar, M.; Sukupolvi, S.; Normark, S.

Mol. Microbiol. 7, 523-536, 1993

A:Title: The RpoS sigma factor relieves H-NS-mediated transcriptional repression of csgA,

A:Reference number: S31202; MUID:93211294; PMID:8459772

A:Accession: S31202

A:Molecule type: DNA

A:Residues: 1-6, 'V', 8-151 <OLS1>

A:Cross-references: EMBL:L04979

A:Accession: S34560

A:Molecule type: protein

A:Residues: 21-42, 44-50 <OLS2>

R:Olsen, A.N.; Arngqvist, A.M.

submitted to the EMBL Data Library, October 1992

A:Reference number: S34559

A:Accession: S34559

A:Molecule type: DNA

A:Residues: 1-133, 'RQRDSGLW' <OLS3>

A:Cross-references: EMBL:L04979; NID:g290424; PIDN:AAA23616.1; PID:g290425

A:Experimental source: strain K-12, substrain W3110

C:Genetics:

A:Gene: csgA

A:Map position: 23.15

C:Function:

A:Description: major component of wild-type curli; interaction between CsgA and CsgB trig and H-kinogen; in the absence of CsgA, CsgB can self-assemble into polymers

F1-20/Domain: signal sequence #status predicted <SIG>

F121-151/Product: curlin #status experimental <MAT>

Query Match 68.8%; Score 533; DB 2; Length 151;

Best Local Similarity 70.2%; Pred. No. 7.1e-38;

Matches 106; Conservative 18; Mismatches 27; Indels 0; Gaps 0;

Qy 1 MKLLKVAAPAAIVVSGSALAGVVPQWGGGNGHNGGNSGPDSTLSIYQYGSANAALALQ 60

Db 1 MKLLKVAAPAAIVVSGSALAGVVPQWGGGNGHNGGNSGPDSTLSIYQYGSANAALALQ 60

Qy 61 SDARKSETTITQSGYNGADVGAGDNSTIELTQNGFRNNATIDQWAKNYDQLVTRVVT 120

Db 61 TDARNSTLTITQHGNGAGDVGGSDSSDITLQRFHGSATLDQWNGKNSNTVQKFGG 120

Qy 121 HEMAHANQTASDSSVMVRQVGFNNATANQY 151

Db 121 GNCAAVDQTASNSVNVTVQFGFNATAHQY 151

RESULT 4

D90806

curlin major subunit CsgA [imported] - Escherichia coli (strain O157:H7, substrain RIMD)

C:Species: Escherichia coli

C:Date: 18-Jul-2001 #sequence\_revision 18-Jul-2001 #text\_change 18-Jul-2001

C:Accession: D90806

R:Havasi, T.; Makino, K.; Ohnishi, M.; Kurokawa, K.; Ishii, K.; Yokoyama, K.; Han, C.G.; Gasakura, N.; Yasunaga, T.; Kuhara, S.; Shiba, T.; Hattori, M.; Shingawa, H.

DNA Res. 8, 11-22, 2001

A:Title: Complete genome sequence of enterohemorrhagic Escherichia coli O157:H7 and genomic

A:Reference number: A99629; MUID:21156231; PMID:11258796

A:Accession: D90806

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-152 <HAY>

A:Cross-references: GB:BA000007; PIDN:BA34843.1; PID:g13360880; GSPDB:GN00154

A:Experimental source: strain O157:H7, substrain RIMD 0509952

C:Genetics:

A:Gene: Ecs1420

Query Match 66.0%; Score 511.5; DB 2; Length 152;

Best Local Similarity 68.4%; Pred. No. 4.6e-36;

Matches 104; Conservative 19; Mismatches 28; Indels 1; Gaps 1;

Qy 1 MKLLKVAAPAAIVVSGSALAGVVPQWGGGNGHNGGNSGPDSTLSIYQYGSANAALAL 59

Db 1 MKLLKVAAPAAIVVSGSALAGVVPQWGGGNGHNGGNSGPDSTLSIYQYGSANAALAL 60

Qy 60 QSDARKSETTITQSGYNGADVGAGDNSTIELTQNGFRNNATIDQWAKNYDQLVTRVVT 119



Db 61 QADARNEDLITQGGGNGADVCGGDDSDIDLTQGFNGSATLDQWNGKDSHMTVKQFG 120  
QY 120 THEMAHANQASDSSVVMVROVGFNNATANQY 151  
Db 121 GGNGAAVDQASSTVNVTQVGFNNATANQY 152  
RESULT 5  
H85665  
hypothetical protein csgA [imported] - Escherichia coli (strain O157:H7, substrain EDL93)  
C:Species: Escherichia coli  
C:Date: 16-Feb-2001 #sequence\_revision 16-Feb-2001 #text\_change 14-Sep-2001  
C:Accession: H85665  
R:Perna, N.T.; Plunkett III, G.; Burland, V.; Mau, B.; Glasner, J.D.; Rose, D.J.; Mayhew  
iller, L.; Grotbeck, E.J.; Davis, N.W.; Lim, A.; Dimalanta, E.; Potamousis, K.; Apodaca,  
Nature 409, 529-533, 2001  
A:Title: Genome sequence of enterohemorrhagic Escherichia coli O157:H7.  
A:Reference number: A85480; MUID:21074935; PMID:11206551  
A:Accession: H85665  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-152 <STO>  
A:Cross-references: GB:AE005174; NID:g12514574; PIDN:AAG55788.1; GSPDB:GN00145; UWGP:Z16  
A:Experimental source: strain O157:H7, substrain EDL933  
C:Genetics:  
A:Gene: csgA  
Query Match 66.0%; Score 511.5; DB 2; Length 152;  
Best Local Similarity 68.4%; Pred. No. 4.6e-36;  
Matches 104; Conservative 19; Mismatches 28; Indels 1; Gaps 1;  
QY 1 MKLLKVAFAAIVSGSALAGVPOW-GGGNGHNGSGGPDSTLSIYQGSANAALAL 59  
Db 1 MKLLKVAFAAIVSGSALAGVPOVGGGNGHNGSGGPDSTLSIYQGSANAALAL 60  
QY 60 QSDARKSETTITQSGYNGADVCGGADNSTIETQNGFRNNATIDQWNAKYDQLVTRV 119  
Db 61 QADARNEDLITQGGGNGADVCGGDDSDIDLTQGFNGSATLDQWNGKDSHMTVKQFG 120  
QY 120 THEMAHANQASDSSVVMVROVGFNNATANQY 151  
Db 121 GGNGAAVDQASSTVNVTQVGFNNATANQY 152  
RESULT 6  
T20847  
hypothetical protein F13E9.4 - Caenorhabditis elegans  
C:Species: Caenorhabditis elegans  
C:Date: 15-Oct-1999 #sequence\_revision 15-Oct-1999 #text\_change 02-Aug-2002  
C:Accession: T20847  
R:McMurray, A.  
submitted to the EMBL Data Library, February 1996  
A:Reference number: Z19332  
A:Accession: T20847  
A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: DNA  
A:Residues: 1-409 <WIL>  
A:Cross-references: EMBL:Z69383; PIDN:CAA93412.1; GSPDB:GN00022; CESP:F13E9.4  
A:Experimental source: clone F13E9  
C:Genetics:  
A:Gene: CESP:F13E9.4  
A:Map position: 4  
A:Introns: 32/1; 275/3; 337/3  
C:Superfamily: loxtrin  
Query Match 13.0%; Score 101; DB 2; Length 409;  
Best Local Similarity 23.3%; Pred. No. 0.4;  
Matches 37; Conservative 26; Mismatches 66; Indels 30; Gaps 5;  
QY 15 SGSALAGVPOVGGGNGHNGSGGPDSTLSIYQGSANAALALQSDARKSETTITQSG 74  
Db 134 SGQASGSMNSFGGCGGQNGQNGFGGCGSFGSGGSGNSLSSANSNGNNQGG--SSSG 191

QY 75 YGN--GADYGGADNSTIE---LTQNGFRNN-----ATIDQWNAK 109  
Db 192 YQNGQRHQCGGCGSHSSSSNSVMSNGYSGNGYNNNGTPTSFLLNVSSSAAQDIYIV 251  
QY 110 NYDQLVTRVVTHEMAHANQASDSSVVMVROVGFNNATAN 148  
Db 252 NKKSLTTQNEQ---ASNWASANSVQAQYIQVETNRS 287  
RESULT 7  
cnjB protein - Tetrahymena thermophila  
C:Species: Tetrahymena thermophila  
C:Date: 19-Mar-1997 #sequence\_revision 18-Jul-1997 #text\_change 07-Dec-1999  
C:Accession: S42136; S42135; S03650  
R:Taylor, F.M.; Martindale, D.W.  
submitted to the EMBL Data Library, October 1992  
A:Reference number: S42136  
A:Accession: S42136  
A:Molecule type: DNA  
A:Residues: 1-1748 <TRAY>  
A:Cross-references: EMBL:L03710; NID:g161751; PID:g161752  
R:Taylor, F.M.; Martindale, D.W.  
Nucleic Acids Res. 21, 4610-4614, 1993  
A:Title: Retroviral-type zinc fingers and glycine-rich repeats in a protein encoded by c  
A:Reference number: S42135; MUID:94051569; PMID:8233798  
A:Accession: S42135  
A:Molecule type: DNA  
A:Residues: 1164-1174; 1179-1198; 1233-1252; 1285-1293; 1297-1309; 1316-1326; 1331-1341; 1343-1  
A:Cross-references: EMBL:L03710  
R:Martindale, D.W.; Taylor, F.M.  
Nucleic Acids Res. 16, 2189-2201, 1988  
A:Title: Multiple introns in a conjugation-specific gene from Tetrahymena thermophila.  
A:Reference number: S03650; MUID:8818981; PMID:3357771  
A:Accession: S03650  
A:Molecule type: DNA  
A:Residues: 236-250; 1', 252-255, 'N', 257-773 <MAR>  
A:Cross-references: EMBL:X06462  
C:Genetics:  
A:Gene: cnjB  
A:Genetic code: SGC5  
A:Introns: 85/3; 136/1; 157/3; 201/2; 290/2; 327/3; 499/1; 573/2; 607/3; 708/3; 777/3; 8  
C:Keywords: zinc finger  
F:1164-1450/Region: glycine-rich  
F:1451-1464/Region: zinc finger CCHC motif  
F:1478-1491/Region: zinc finger CCHC motif  
F:1501-1514/Region: zinc finger CCHC motif  
F:1530-1543/Region: zinc finger CCHC motif  
F:1555-1568/Region: zinc finger CCHC motif  
F:1579-1592/Region: zinc finger CCHC motif  
F:1602-1615/Region: zinc finger CCHC motif  
F:1626-1748/Region: glycine-rich  
Query Match 12.8%; Score 99; DB 2; Length 1748;  
Best Local Similarity 28.0%; Pred. No. 3.1;  
Matches 37; Conservative 16; Mismatches 31; Indels 48; Gaps 6;  
QY 25 QWGGGNGHNGG---GNSGPDSTLSIYQGSANAALQSDARKSETTIT---QSGYGN 77  
Db 1640 QFGGNGSGGSGWGTSSTSSGDMN-----COSNVQESTTSSGGWGGSGGN 1685  
QY 78 GADVGGADNSTIETQNGFRNNATIDQWNAKYDQLVTRVVTHEMAHANQASDSSVMV 137  
Db 1686 QTGGGWSNDN-----QQQNTGGGWSGN-----SNQTNSS--- 1722  
QY 138 QVGFNNATAN 149  
Db 1723 ---WGSNNQAS 1730  
RESULT 8  
JC6040

fimbrin protein agfB precursor - Salmonella enteritidis  
 C/Species: Salmonella enteritidis  
 C/Date: 31-Dec-1996 #sequence\_revision 31-Dec-1996 #text\_change 08-Oct-1999  
 C/Accession: J06040  
 R/Collinson, S.K.; Cloughier, S.C.; Doran, J.L.; Baner, P.A.; Kay, W.W.  
 J. Bacteriol. 178, 662-667, 1996  
 A/Title: Salmonella enteritidis agfBAC operon encoding thin, aggregative fimbriae.  
 A/Reference number: J06039; MUID:96146512; PMID:8550497  
 A/Accession: J06040  
 A/Molecule type: DNA  
 A/Residues: 1-151 <COL>  
 A/Cross-references: GB:U43280; NID:gl184712; PIDN:AA043598.1; PID:gl184713  
 A/Experimental source: strain 276755-3b  
 C/Genetics:  
 A/Gene: agfB  
 A/Description: minor component of thin aggregative fimbriae  
 A/Note: fimbriae bind to fibronectin, plasminogen, tissue plasminogen activator  
 C/Keywords: fimbria  
 F:1-21/Domain: signal sequence #status predicted <SIG>  
 F:22-151/Product: fimbrin protein agfB #status predicted <MAT>

Query Match 12.7%; Score 98.5; DB 2; Length 151;  
 Best Local Similarity 28.8%; Pred. No. 0.21;  
 Matches 34; Conservative 17; Mismatches 48; Indels 19; Gaps 5;

QY 51 GSANALALQSDARKSE-----TTTQSGYNGADVQ-GADNST-----IELTQ 94  
 DB 14 GAPGIATATNYDLARSEYFAVNELSKSFNQAAIIGQVGTDSARVRQEGSKLLSVISQ 73

QY 95 NGRFNATIDQNAKNDY-QLVTRVVTHEMAHANOASDSSVMVQVGFNNATANOY 151  
 DB 74 EGGNFAKVDQ--AGNYNFAIEQTGNANDASISQAYGNSAAITQSGNKNITQY 129

RESULT 9  
 AH0635  
 nucleation component of curlin monomers [imported] - Salmonella enterica subsp. enterica  
 C/Species: Salmonella enterica subsp. enterica serovar Typhi  
 A/Note: this species has also been called Salmonella typhi  
 C/Date: 09-Nov-2001 #sequence\_revision 09-Nov-2001 #text\_change 18-Nov-2002  
 R/Accession: AH0635  
 R/Farkhill, J.; Dougan, G.; James, K.D.; Thomson, N.R.; Pickard, D.; Wain, J.; Churcher, T.; Connor, P.; Cronin, A.; Davis, P.; Davies, R.M.; Dowd, L.; White, N.; Farrar, S.; Moule, S.; O'Gaora, P.  
 Nature 413, 848-852, 2001  
 A/Authors: Parry, C.; Quail, M.; Rutherford, K.; Simmonds, M.; Skelton, J.; Stevens, K.;  
 A/Title: Complete genome sequence of a multiple drug resistant Salmonella enterica serovar AB0502; MUID:21534947; PMID:11677608  
 A/Accession: AH0635  
 A/Status: preliminary  
 A/Molecule type: DNA  
 A/Residues: 1-151 <PAR>  
 A/Cross-references: GB:AL513382; PIDN:CAD08267.1; PID:gl6502314; GSPDB:GN00176  
 C/Genetics:  
 A/Gene: STY1180

Query Match 12.7%; Score 98.5; DB 2; Length 151;  
 Best Local Similarity 28.8%; Pred. No. 0.21;  
 Matches 34; Conservative 17; Mismatches 48; Indels 19; Gaps 5;

QY 51 GSANALALQSDARKSE-----TTTQSGYNGADVQ-GADNST-----IELTQ 94  
 DB 14 GAPGIATATNYDLARSEYFAVNELSKSFNQAAIIGQVGTDSARVRQEGSKLLSVISQ 73

QY 95 NGRFNATIDQNAKNDY-QLVTRVVTHEMAHANOASDSSVMVQVGFNNATANOY 151  
 DB 74 EGGNFAKVDQ--AGNYNFAIEQTGNANDASISQAYGNSAAITQSGNKNITQY 129

RESULT 10  
 S70787  
 curlin nucleator protein csgB precursor - Escherichia coli (strain K-12)

N/Alternate names: csgB protein; curlin nucleation component; minor curlin protein  
 C/Species: Escherichia coli  
 C/Date: 12-Feb-1998 #sequence\_revision 20-Feb-1998 #text\_change 01-Mar-2002  
 C/Accession: S70787; F64846  
 R/Hammar, M.; Arngqvist, A.; Bian, Z.; Olsen, A.; Normark, S.  
 Mol. Microbiol. 18, 661-670, 1995  
 A/Title: Expression of two csg operons is required for production of fibronectin- and Cor  
 A/Reference number: S70783; MUID:96414468; PMID:8817489  
 A/Accession: S70787  
 A/Status: nucleic acid sequence not shown; translation not shown  
 A/Molecule type: DNA  
 A/Residues: 1-151 <HAM>  
 A/Cross-references: EMBL:X0754; NID:gl147558; PIDN:CAA63281.1; PID:gl147563  
 A/Experimental source: strain K12, substrain W3110  
 A/Note: the nucleotide sequence was submitted to the EMBL Data Library, August 1995  
 R/Blattner, F.R.; Plunkett III, G.; Bloch, C.A.; Perna, N.T.; Burland, V.; Riley, M.; Col  
 A.; Rose, D.J.; Mau, B.; Shao, Y.  
 Science 277, 1453-1462, 1997  
 A/Title: The complete genome sequence of Escherichia coli K-12.  
 A/Reference number: A64720; MUID:97426617; PMID:9278503  
 A/Accession: F64846  
 A/Status: nucleic acid sequence not shown; translation not shown  
 A/Molecule type: DNA  
 A/Residues: 1-151 <BLAT>  
 A/Cross-references: GB:AE000205; GB:U00096; NID:gl787265; PIDN:AA074125.1; PID:gl787278;  
 A/Experimental source: strain K-12, substrain MG1655  
 C/Genetics:  
 A/Gene: csgB  
 A/Map position: 23.15  
 C/Function:  
 A/Description: minor component of wild-type curli; interaction between CsgA and CsgB trig  
 A/Note: curli are thin, coiled fibers expressed on the surface of Escherichia coli that i  
 and H-kininogen; in the absence of CsgA, CsgB can self-assemble into polymers  
 F:1-21/Domain: signal sequence #status predicted <SIG>  
 F:22-151/Product: minor curlin chain #status predicted <MAT>

Query Match 12.5%; Score 96.5; DB 2; Length 151;  
 Best Local Similarity 25.6%; Pred. No. 0.31;  
 Matches 30; Conservative 24; Mismatches 46; Indels 17; Gaps 4;

QY 51 GSANALALQSDARKSE-----TTTQSGYNGADVQGGADNSTIELTQGFNNATIDQ 105  
 DB 14 GAPGIAAAGYDLANGSEYFAVNELSKSFNQAAIIGQAGTNSAQLRQGGSKLLAVAAQ 73

QY 106 WNAKNYDQVTRVVTHEMAHANO--TASDSSV-----WVQVGFNNATANOY 151  
 DB 74 EGSSNRKAK-IDQTDYNYLAYIDQGSANDASISQAYGNTAMIIQKSGNKNITQY 129

RESULT 11  
 C90806  
 minor curlin subunit precursor CsgB [imported] - Escherichia coli (strain O157:H7, substr  
 C/Species: Escherichia coli  
 C/Date: 18-Jul-2001 #sequence\_revision 18-Jul-2001 #text\_change 18-Jul-2001  
 C/Accession: C90806  
 R/Hayaishi, T.; Makino, K.; Ohnishi, M.; Kurokawa, K.; Ishii, K.; Yokoyama, K.; Han, C.G.;  
 gasawara, N.; Yasunaga, T.; Kuhara, S.; Shiba, T.; Hattori, M.; Shinagawa, H.  
 DNA Res. 8, 11-22, 2001  
 A/Title: Complete genome sequence of enterohemorrhagic Escherichia coli O157:H7 and genom  
 A/Reference number: A99629; MUID:21156231; PMID:11258796  
 A/Accession: C90806  
 A/Status: preliminary  
 A/Molecule type: DNA  
 A/Residues: 1-151 <HAY>  
 A/Cross-references: GB:BA000007; PIDN:BA034842.1; PID:gl3360879; GSPDB:GN00154  
 A/Experimental source: strain O157:H7, substrain RIMD 0509952  
 C/Genetics:  
 A/Gene: Ecs1419

Query Match 12.5%; Score 96.5; DB 2; Length 151;  
 Best Local Similarity 25.6%; Pred. No. 0.31;  
 Matches 30; Conservative 24; Mismatches 46; Indels 17; Gaps 4;

Query March	11.9%	Score 92.5	DB 2	Length 1028
Best Local Similarity	26.7%	Prod No. 5	9	
Marches 40	Conservative 14	Mismatches 61	Indels 35	Gaps 6
3 ILKVAEFAATVVGSSALAGVVPDWGGGHNHGGGNSGGPDSTLSIVYCGSANAALQSD	62			

Db	59	LQNAAYIMSAGSG-----GGCTGNGGGGASGPGGSPANSGGGGGG-----	104
Qy	63	ARKSETTITOSGYGNGADVGGADNSTIELTQNGFRNNATIDQWNAKNDQLVTRVWTHE	122
Db	105	-----GGGYINGGVG-GPNS--LDGNNLNFASVSNYESN-----SKFHNH	147
Qy	123	MAHA-----NOTASDSSVMVRQVFGNNATA	148
Db	148	HHQHNNNNNNNGQTSMGMHPYGGNPSA	177

Search completed: August 2, 2004, 14:56:24  
Job time : 9.4 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: August 2, 2004, 14:36:12 ; Search time 5.3 Seconds  
(without alignments)  
1483.508 Million cell updates/sec

Title: US-09-543-407-24  
Perfect score: 775  
Sequence: 1 MKLLKVAFAAIVVSGSALA.....DSSVMVRQVGFNNATANQY 151

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 141681 seqs, 52070155 residues  
Total number of hits satisfying chosen parameters: 141681

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : SwissProt\_42.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	690	89.0	151	1	CSGA_SALTY
2	533	68.8	151	1	CSGA_ECOLI
3	511.5	66.0	152	1	CSGA_ECO57
4	98.5	12.7	151	1	CSGB_SALTY
5	98.5	12.7	151	1	CSGB_ECOLI
6	96.5	12.5	151	1	CSGB_SALTY
7	92.5	11.9	1028	1	OVO_DROME
8	90.5	11.7	1327	1	TNKI_HUMAN
9	90	11.6	347	1	MSA2_PLAF2
10	89	11.5	262	1	VG38_BPT2
11	87	11.2	590	1	GP63_LEIDO
12	87	11.2	599	1	GP63_LEICH
13	87	11.2	602	1	GP63_LEIMA
14	87	11.2	646	1	GP63_LEIME
15	87	11.2	1656	1	OMP8_RICJA
16	85	11.0	401	1	YK03_CABEL
17	82	10.6	678	1	YF48_MYCTU
18	82	10.6	1093	1	PER_DROWI
19	82	10.6	1115	1	TBC2_CHLRE
20	81.5	10.5	306	1	HMUX_DROPS
21	81.5	10.5	311	1	YB4C_PSSPP
22	81.5	10.5	392	1	HME1_HUMAN
23	81	10.5	165	1	GRP1_ORYSA
24	80.5	10.4	548	1	CEAK_ECOLI
25	80.5	10.4	720	1	G7AC_BREDI
26	80.5	10.4	894	1	ILF3_HUMAN
27	80	10.3	427	1	CF1A_DROME
28	80	10.3	491	1	YK98_MYCTU
29	79.5	10.3	172	1	CH18_DROME
30	79.5	10.3	760	1	YB1L_ECOLI
31	79.5	10.3	1034	1	ICEN_PANAN
32	79.5	10.3	1258	1	ICEN_ERWHE
33	79.5	10.3	1322	1	ICEA_PANAN

ALIGNMENTS

RESULT 1

ID	CSGA_SALTY	STANDARD;	PRT;	151 AA.
AC	P55225;			
DT	01-OCT-1996 (Rel. 34, Created)			
DT	01-OCT-1996 (Rel. 34, Last sequence update)			
DT	10-OCT-2003 (Rel. 42, Last annotation update)			
DE	Major curlin subunit precursor (Fimbrin SEF17).			
GN	CSGA OR AGFA OR STM1144 OR STY1181 OR T1776.			
OS	Salmonella typhimurium,			
OS	Salmonella typhi, and			
OS	Salmonella enteritidis.			
OC	Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;			
OC	Enterobacteriaceae; Salmonella.			
OX	NCBI_TaxID=602, 601, 592;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RC	SPECIES=S.typhimurium; STRAIN=SR-11;			
RX	MEDLINE=98117058; PubMed=9457880;			
RA	Romling U., Bian Z., Hammar M., Sierralta W.D., Normark S.;			
RA	"Curli fibers are highly conserved between Salmonella typhimurium and			
RT	Escherichia coli with respect to operon structure and regulation.";			
RL	J. Bacteriol. 180:722-731(1998).			
RN	[2]			
RP	SEQUENCE FROM N.A.			
RC	SPECIES=S.typhimurium; STRAIN=LT2 / SGSC1412 / ATCC 700720;			
RX	MEDLINE=21534948; PubMed=11677609;			
RA	McClelland M., Sanderson K.E., Spieth J., Clifton S.W., Latreille P.,			
RA	Courtney L., Porwollik S., Ali J., Dente M., Du F., Hou S., Layman D.,			
RA	Leonard S., Nguyen C., Scott K., Holmes A., Grewal N., Mulvaney E.,			
RA	Ryan E., Sun H., Florea L., Miller W., Stoneking T., Nhan M.,			
RA	Waterston R., Wilson R.K.;			
RT	"Complete genome sequence of Salmonella enterica serovar Typhimurium			
RL	LT2.";			
RL	Nature 413:852-856(2001).			
RN	[3]			
RP	SEQUENCE FROM N.A.			
RC	SPECIES=S.typhi; STRAIN=CT18;			
RX	MEDLINE=21534947; PubMed=11677608;			
RA	Parkhill J., Dougan G., James K.D., Thomson N.R., Pickard D., Main J.,			
RA	Churcher C., Mungall K.L., Bentley S.D., Holden M.T.G., Sebahia M.,			
RA	Baker S., Basham P., Brooks K., Chillingworth T., Connerton P.,			
RA	Cronin A., Davis P., Davies R.M., Dowd L., White N., Farrar J.,			
RA	Feltwell T., Hamlin N., Haque A., Hien T.T., Holroyd S., Jagels K.,			
RA	Krogh A., Larsen T.S., Leather S., Moule S., O'Gaora P., Parry C.,			
RA	Quail M.A., Rutherford K., Simmonds M., Skelton J., Stevens K.,			
RA	Whitehead S., Barrall B.G.;			
RT	"Complete genome sequence of a multiple drug resistant Salmonella			
RL	enterica serovar Typhi CT18.";			
RL	Nature 413:848-852(2001).			
RN	[4]			
RP	SEQUENCE FROM N.A.			
RC	SPECIES=S.typhi; STRAIN=Ty2 / ATCC 700931;			
RX	MEDLINE=22531367; PubMed=12644504;			
RA	Deng W., Liou S.-R., Plunkett G. III, Mayhew G.F., Rose D.J.,			
RA	Burland V., Kodoyianni V., Schwartz D.C., Blattner F.R.;			

P30130	escherichia
Q81xf0	homo sapien
Q9zky0	helicobacte
P50610	helicobacte
P18431	drosophila
P52302	drosophila
P38742	saccharomyc
Q12756	homo sapien
O66408	aquifex aeo
P22670	homo sapien
Q33479	pseudomonas
P09815	pseudomonas

RT "Comparative genomics of *Salmonella enterica* serovar Typhi strains Ty2  
 RT and CT18.";  
 RT J. Bacteriol. 185:2330-2337(2003).  
 RN [5]  
 RP SEQUENCE FROM N.A.  
 RC SPECIES=S. enteritidis; STRAIN=27655-3B;  
 RX MEDLINE=96146512; PubMed=8550497;  
 RA Collinson S.K., Clouthier S.C., Doran J.L., Baner P.A., Kay W.W.;  
 RT "Salmonella enteritidis agfBAC operon encoding thin, aggregative  
 RT fimbriae.";  
 RT J. Bacteriol. 178:662-667(1996).  
 RN [6]  
 RP SEQUENCE OF 21-151 FROM N.A.  
 RC SPECIES=S. enteritidis; STRAIN=27655-3B;  
 RX MEDLINE=94013373; PubMed=8104959;  
 RA Doran J.L., Collinson S.K., Burian J., Sarlos G., Todd B.C.D.,  
 Munro C.K., Kay C.M., Baner P.A., Peterkin P.I., Kay W.W.;  
 RT "DNA-based diagnostic tests for *Salmonella* species targeting agfA,  
 RT the structural gene for thin, aggregative fimbriae.";  
 RL J. Clin. Microbiol. 31:2263-2273(1993).  
 RN [7]  
 RP SEQUENCE OF 21-33.  
 RC SPECIES=S. enteritidis; STRAIN=27655-3B;  
 RX MEDLINE=91310586; PubMed=1677357;  
 RA Collinson S.K., Emeody L., Mueller K.-M., Trust T.J., Kay W.W.;  
 RT "Purification and characterization of thin, aggregative fimbriae from  
 RT *Salmonella enteritidis*.";  
 RL J. Bacteriol. 173:4773-4781(1991).  
 CC -1- FUNCTION: CURLIN IS THE STRUCTURAL SUBUNIT OF THE CURLI. CURLI ARE  
 CC COILED SURFACE STRUCTURES THAT ASSEMBLE PREFERENTIALLY AT GROWTH  
 CC TEMPERATURES BELOW 37 DEGREES CELSIUS. CURLI CAN BIND TO  
 CC FIBRONECTIN.  
 CC -1- SIMILARITY: BELONGS TO THE CSGA/CSGB FAMILY.  
 CC -----  
 CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL; AJ002301; CAA05317.1; -;  
 DR EMBL; AE008749; AAL20074.1; -;  
 DR EMBL; AL627269; CAD08268.1; -;  
 DR EMBL; AE016940; AAC69399.1; -;  
 DR EMBL; U43280; AAC43599.1; -;  
 DR PIR; JC6039; JC6039;  
 DR StyGene; SG10608; csGA.  
 KW Fimbria; Signal; Complete proteome.  
 FT SIGNAL 1 20  
 FT CHAIN 21 151 MAJOR CURLIN SUBUNIT.  
 FT CONFLICT 134 151 SYMVQRQVGGNNATANYQ -> DSYTQVAS (IN  
 FT RES. 6).  
 FT SEQUENCE 151 AA; 15305 MW; B7DAC0D16B621359 CRC64;  
 Query Match 89.0%; Score 690; DB 1; Length 151;  
 Best Local Similarity 91.4%; Pred. No. 1.4e-51;  
 Matches 138; Conservative 2; Mismatches 11; Indels 0; Gaps 0;  
 QY 1 MKLLKVAFAAIVVSGSALAGVVPQWGGGNGGNGSGSPDSTLSIYQGSANAALALQ 60  
 Db 1 MKLLKVAFAAIVVSGSALAGVVPQWGGGNGGNGSGSPDSTLSIYQGSANAALALQ 60  
 QY 61 SPARKSETTITSGYNGADVCGQADNSTIETQNGFRNNATIDOWNAKNVQDLVTRVVT 120  
 Db 61 SPARKSETTITSGYNGADVCGQADNSTIETQNGFRNNATIDOWNAKNVQDLVTRVVT 120  
 QY 121 HEMAHANQTSDDSVVVRQVGGNNATANYQ 151  
 Db 121 NNAALVNQTSDDSVVVRQVGGNNATANYQ 151

RESULT 2  
 CSGA\_ECOLI  
 ID CSGA\_ECOLI STANDARD; PRT; 151 AA.  
 AC P28307;  
 DT 01-DEC-1992 (Rel. 24, Created)  
 DT 01-OCT-1996 (Rel. 34, Last sequence update)  
 DT 28-FEB-2003 (Rel. 41, Last annotation update)  
 DE Major curlin subunit precursor.  
 GN CSGA OR B1042.  
 OS *Escherichia coli*.  
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;  
 OC Enterobacteriaceae; *Escherichia*.  
 OX NCBI\_TaxID=562;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=K12 / W3110;  
 RX MEDLINE=93211294; PubMed=8459772;  
 RA Olsen A., Arngqvist A.;  
 RT "The RpoS sigma factor relieves H-NS-mediated transcriptional  
 RT repression of csGA, the subunit gene of fibronectin-binding curli in  
 RT *Escherichia coli*.";  
 RL Mol. Microbiol. 7:523-536(1993).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=K12 / MC4100;  
 RX MEDLINE=96414468; PubMed=8817489;  
 RA Hammar M., Arngqvist A., Bian Z., Olsen A., Normark S.;  
 RT "Expression of two csG operons is required for production of  
 RT fibronectin- and congo red-binding curli polymers in *Escherichia coli*  
 RT K-12.";  
 RL Mol. Microbiol. 18:661-670(1995).  
 RN [3]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=K12 / MG1655;  
 RX MEDLINE=97426617; PubMed=9278503;  
 RA Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,  
 Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,  
 Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,  
 Mau B., Shao Y.;  
 RT "The complete genome sequence of *Escherichia coli* K-12.";  
 RL Science 277:1234-1238(1997).  
 RN [4]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=K12;  
 RX MEDLINE=97061202; PubMed=8905232;  
 RA Oshima T., Aiba H., Baba T., Fujita K., Hayaashi K., Honjo A.,  
 Ikemoto K., Inada T., Itoh T., Kajihara M., Kanai K., Kashimoto K.,  
 Kimura S., Kitagawa M., Makino K., Masuda S., Miki T., Mizobuchi K.,  
 Mori H., Motomura K., Nakamura Y., Nishimoto H., Nishio Y., Saito N.,  
 Sempel G., Seki Y., Tagami H., Takemoto K., Wada C., Yamamoto Y.,  
 Yano M., Horuchi T.;  
 RT "A 718-kb DNA sequence of the *Escherichia coli* K-12 genome  
 RT corresponding to the 12.7-28.0 min region on the linkage map.";  
 RL DNA Res. 3:137-155(1996).  
 RN [5]  
 RP SEQUENCE OF 21-40.  
 RC STRAIN=K12 / YMEI;  
 RX MEDLINE=93023873; PubMed=1357528;  
 RA Arngqvist A., Olsen A., Pfeifer J., Russell D.G., Normark S.;  
 RT "The Cur protein activates cryptic genes for curli formation and  
 RT fibronectin binding in *Escherichia coli* HB101.";  
 RL Mol. Microbiol. 6:2443-2452(1992).  
 RN [6]  
 RP SEQUENCE OF 21-31.  
 RX MEDLINE=91310586; PubMed=1677357;  
 RA Collinson S.K., Emeody L., Trust T.J., Kay W.W.;  
 RT "Purification and characterization of thin, aggregative fimbriae from  
 RT *Salmonella enteritidis*.";  
 RL J. Bacteriol. 173:4773-4781(1991).  
 CC -1- FUNCTION: CURLIN IS THE STRUCTURAL SUBUNIT OF THE CURLI. CURLI ARE  
 CC COILED SURFACE STRUCTURES THAT ASSEMBLE PREFERENTIALLY AT GROWTH  
 CC TEMPERATURES BELOW 37 DEGREES CELSIUS. CURLI CAN BIND TO  
 CC FIBRONECTIN.  
 CC -----

```
[3]
SEQUENCE FROM N.A.
STRN=O157:H7 / RIMD 0509952;
MEDLINE=21156231; PubMed=11258796;
HAYASHI T., Makino K., Ohnishi M., Kurokawa K., Ishii K., Yokoyama K.,
HAN C.-T., Ohtsubo E., Nakayama K., Murata T., Tanaka M., Tohe T.,
RAIDA T., Takami H., Honda T., Sasakawa C., Ogasawara N., Yasunaga T.,
Kuhara S., Shiba T., Hattori M., Shinagawa H.,
"Complete genome sequence of enterohemorrhagic Escherichia coli
O157:H7 and genomic comparison with a laboratory strain K-12."
DNA RES. 8:11-22(2001).
CC -!- FUNCTION: CURLIN IS THE STRUCTURAL SUBUNIT OF THE CURLI. CURLI ARE
CC COILED SURFACE STRUCTURES THAT ASSEMBLE PREFERENTIALLY AT GROWTH
CC TEMPERATURES BELOW 37 DEGREES CELSIUS. CURLI CAN BIND TO
CC FIBROBLASTS.
CC -!- SIMILARITY: BELONGS TO THE CSGA/CSGB FAMILY.
CC
CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation
CC at the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch)
CC -----
CC EMBL; AF275733; AAK53212.1; --
CC EMBL; AE005315; AAG55788.1; --
CC EMBL; AP002854; BAB34843.1; --
CC FIR; D90806; D90806.
CC FIR; H85665; H85665.
CC FIRBRIA; Signal; Complete proteome.
CC SIGNAL 1 20 BY SIMILARITY.
CC CHAIN 21 152 MAJOR CURLIN SUBUNIT.
CC SEQUENCE 152 AA; 15099 MW; EE2D2D94DDE91243 CRC64;
CC
Query Match 56.0%; Score 511.5; DB 1; Length 152;
Best Local Similarity 58.4%; Pred.No.1.5e-36; Indels 1; Gaps 1;
Matches 104; Conservative 19; Mismatches 28;
QY 1 MKLLKVAFAAAIVVSGSALAGVVPQW-GGGGNHNGGKNSGSPDSTLSIYQYGSAALAL 59
DB 1 MKLLKVAIAAI VFGSSALAGVVPQYGGGGNHGGGNGSNPSELNIYQYGGNSALAL 60
QY 60 QSDARKSETTITQSGYNGADVGCGADNSTIELTQTNGFRNNATIDWNKAKYDQLVTRVV 119
DB 61 QADARNSDLAITQHGCGANGADVQGGSDDSIDLTQRGFNGSATLDQWNGKDSHMTVKQF 120
QY 120 THEMAHQNTASDVSVYVRQVGGNNATANQY 151
DB 121 GCNGAANDQTASNTVNVTQVGGNNATAHQY 152
RESULT 4
CSGB SALT1 STANDARD; PRT; 151 AA.
AC Q8ZTN3;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Minor curlin subunit precursor.
GN CSGB OR STYL180 OR TL777.
OS Salmonella typhi
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Salmonella.
OX NCBI_TaxID=601;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=CT18;
RX MEDLINE=21534947; PubMed=11677608;
RA Parkhill J., Dougan G., James K.D., Thomson N.R., Pickard D., Wain J.,
RA Churcher C., Mungall K.L., Bentley S.D., Holden M.T.G., Sebaihia M.,
BAKER S., Basham D., Brooks K., Chillingworth T., Cornerton P.,
RA Cronin A., Davis P., Davies R.M., Dowd L., White N., Farrar
```





RP SEQUENCE FROM N.A.  
RC STRAIN=K12 / MC4100;  
RX MEDLINE=96414468; PubMed=8917489;  
RA Hammar M., Arqvist A., Bian Z., Olsen A., Normark S.;  
RT "Expression of two csg operons is required for production of  
fibronectin- and Congo red-binding curli polymers in *Escherichia coli*  
K-12";  
RL Mol. Microbiol. 18:661-670 (1998).  
RN [2]  
RP SEQUENCE FROM N.A.  
RC STRAIN=K12 / MG1655;  
RX MEDLINE=97426617; PubMed=9278503;  
RA Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,  
RW Bacterial F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,  
RA Rileigh M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,  
RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,  
RA Mau B., Shao Y.;  
RT "The complete genome sequence of *Escherichia coli* K-12";  
RL Science 277:1453-1474 (1997).  
RN [3]  
RP SEQUENCE FROM N.A.  
RC STRAIN=K12;  
RX MEDLINE=97061202; PubMed=8905232;  
RA Oshima T., Alba H., Baba T., Fujita K., Hayashi K., Honjo A.,  
RA Ikemoto K., Inada T., Itoh T., Kajihara M., Kanai K., Kashimoto K.,  
RA Kimura S., Kitagawa M., Makino K., Masuda S., Miki T., Mizobuchi K.,  
RA Mori H., Motomura K., Nakamura Y., Nishimoto H., Nishio Y., Saito N.,  
RA Sampei G., Seki Y., Tagami H., Takemoto K., Wada C., Yamamoto Y.,  
RA Yano M., Horiuchi T.;  
RT "A 718-kb DNA sequence of the *Escherichia coli* K-12 genome  
corresponding to the 12.7-28.0 min region on the linkage map";  
RL DNA Res. 3:137-155 (1996).  
RN [4]  
RP SEQUENCE FROM N.A.  
RC STRAIN=0157:H7 / EDL933 / ATCC 700927;  
RX MEDLINE=21074935; PubMed=11206551;  
RA Perna N.T., Plunkett G. III, Burland V., Mau B., Glasner J.D.,  
RA Rose D.J., Mayhew G.F., Evans P.S., Gregor J., Kirkpatrick H.A.,  
RA Postfai G., Hackett J., Klink S., Boutin A., Shao Y., Miller L.,  
RA Grobbeck E.J., Davis N.W., Lim A., Dimalanta E.T., Potamousis K.,  
RA Apodaca J., Anantharaman T.S., Lin J., Yen G., Schwartz D.C.,  
RA Welch R.A., Blattner F.R.;  
RT "Genome sequence of enterohaemorrhagic *Escherichia coli* O157:H7";  
RL Nature 409:529-533 (2001).  
RN [5]  
RP SEQUENCE FROM N.A.  
RC STRAIN=0157:H7 / RIMD 0509952;  
RX MEDLINE=21156231; PubMed=11258796;  
RA Hayashi T., Makino K., Ohnishi M., Kurokawa K., Ishii K., Yokoyama K.,  
RA Han C.-G., Ohtsubo E., Nakayama K., Murata T., Tanaka M., Tobe T.,  
RA Iida T., Takami H., Honda T., Sasakawa C., Ogasawara N., Yasunaga T.,  
RA Kuhara S., Shiba T., Hattori M., Shingawa H.;  
RT "Complete genome sequence of enterohaemorrhagic *Escherichia coli*  
O157:H7 and genomic comparison with a laboratory strain K-12";  
RL DNA Res. 8:11-22 (2001).  
RN [6]  
RP SEQUENCE OF 1-21 FROM N.A.  
RC STRAIN=K12;  
RX MEDLINE=95157246; PubMed=7854117;  
RA Arqvist A., Olsen A., Normark S.;  
RT "Sigma S-dependent growth-phase induction of the *csgBA* promoter in  
*Escherichia coli* can be achieved in vivo by sigma 70 in the absence  
of the nucleoid-associated protein H-NS";  
RL Mol. Microbiol. 13:1021-1032 (1994).  
CC -!- FUNCTION: CURLIN IS THE STRUCTURAL SUBUNIT OF THE CURLI. CURLI ARE  
COILED SURFACE STRUCTURES THAT ASSEMBLE PREFERENTIALLY AT GROWTH  
TEMPERATURES BELOW 37 DEGREES CELSIUS. CURLI CAN BIND TO  
FIBRONECTIN. THE MINOR SUBUNIT IS THE NUCLEATION COMPONENT OF  
CURLIN MONOMERS.  
CC -!- SIMILARITY: BELONGS TO THE CSGA/CSGB FAMILY.  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
between the Swiss Institute of Bioinformatics and the EMBL outstation -  
the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
CC EMBL; X90754; CAA62281.1; -;  
DR EMBL; AE000205; AAC74125.1; -;  
DR EMBL; D90741; BAA35831.1; -;  
DR EMBL; AE005315; AAG55787.1; -;  
DR EMBL; AP002554; BAB34842.1; -;  
DR PIR; C90806; C90806;  
DR PIR; G85665; G85665;  
DR PIR; S70787; S70787;  
DR ECoGene; EG12621; csgB.  
KW Fibria; Signal; Complete proteome.  
FT CHAIN 1 21 POTENTIAL.  
FT CHAIN 22 151 MINOR CURLIN SUBUNIT.  
SQ SEQUENCE 151 AA; 15882 MW; B18D266B964014B8 CRC64;  
  
Query Match 12.5%; Score 96.5; DB 1; Length 151;  
Best Local Similarity 25.6%; Pred. No. 0.15;  
Matches 30; Conservative 24; Mismatches 46; Indels 17; Gaps 4;  
  
QY 51 GSANAALQSDARKS-----TTITQSGYGNQADVGQADNSTIELTQNGFRNNATIDQ 105  
Db 14 GAPGIAAGVLDLANSEYNFAVNELSKSFNQAIIGQATNNNSAQLRQGSKLLAVVAQ 73  
QY 106 WNAKXYDQLVTRVVTHEMAHANQ--TASDSSV-----MVRQVGFNNATANQY 151  
Db 74 EGSSNRK-IDQTGDYNLAVIDQAGSANDASISQAYGNTAMTIQSGNKNANTQY 129  
  
RESULT 7  
OVO DROME  
ID - OVO DROME STANDARD; PRT; 1028 AA.  
AC P51521; Q9XZU4;  
DT 01-OCT-1996 (Rel. 34, Created)  
DT 01-OCT-1996 (Rel. 34, Last sequence update)  
DT 10-OCT-2003 (Rel. 42, Last annotation update)  
DE OVO protein (Shaven baby protein).  
GN OVO OR SVE.  
OS Drosophila melanogaster (Fruit fly).  
OC Eukaryota; Metazoa; Arthropoda; Insecta; Pterygota;  
OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;  
OC Ephydroidea; Drosophilidae; Drosophila.  
CX NCBI\_TaxID=7227;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC TISSUE=Ovary;  
RX MEDLINE=95021209; PubMed=7935398;  
RA Garfinkel M.D., Wang J., Liang Y., Mahowald A.P.;  
RT "Multiple products from the shavenbaby-ovo gene region of *Drosophila*  
melanogaster: relationship to genetic complexity";  
RL Mol. Cell. Biol. 14:6809-6818 (1994).  
RN [2]  
RP SEQUENCE FROM N.A.  
RC STRAIN=Oregon-R;  
RX MEDLINE=91293102; PubMed=1712294;  
RA Mevel-Ninio M.T.M., Terracol R., Kafatos F.C.;  
RT "The ovo gene of *Drosophila* encodes a zinc finger protein required  
for female germ line development";  
RL EMO J. 10:2259-2266 (1991).  
CC -!- FUNCTION: REQUIRED FOR SURVIVAL AND DIFFERENTIATION OF FEMALE GERM  
LINE CELLS. PLAYS A ROLE IN GERM LINE SEX DETERMINATION.  
CC -!- SUBCELLULAR LOCATION: Nuclear (Potential).  
CC -!- DEVELOPMENTAL STAGE: FIRST APPEARS IN THE GERMARIUM AND  
ACCUMULATES IN NURSE CELLS DURING OÖGENESIS. STORED IN THE EGG,  
BUT IS RAPIDLY LOST IN THE EMBRYO EXCEPT FOR ITS CONTINUED  
PRESENCE IN THE GERM LINE PRECURSOR POLE CELLS.  
CC -!- SIMILARITY: Contains 4 C2H2-type zinc fingers.  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
between the Swiss Institute of Bioinformatics and the EMBL outstation -

CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----

CC EMBL; U11383; AAB60216.1; -  
 DR EMBL; X59772; CAB36921.1; ALT\_SEQ.  
 DR PIR; A56038; A56038.  
 DR HSP; P07248; ZADR.  
 DR TRANSFAC; T00669; -  
 DR FlyBase; FBgn003028; ovo.  
 DR InterPro; IPR007087; Znf.C2H2.  
 DR Pfam; PF00096; Zf-C2H2; 3.  
 DR SMART; SM00355; Znf.C2H2; 4.  
 DR PROSITE; PS00028; ZINC\_FINGER\_C2H2\_1; 3.  
 DR PROSITE; PS00157; ZINC\_FINGER\_C2H2\_2; 3.  
 KW Zinc-finger; Metal-binding; DNA-binding; Repeat; Nuclear protein;  
 KW Transcription regulation.  
 FT DOMAIN 62 66 POLY-ALA.  
 FT DOMAIN 72 77 POLY-GLY.  
 FT DOMAIN 80 85 POLY-GLY.  
 FT DOMAIN 98 108 POLY-GLY.  
 FT DOMAIN 144 152 POLY-HIS.  
 FT DOMAIN 153 159 POLY-ASN.  
 FT DOMAIN 336 339 POLY-GLN.  
 FT DOMAIN 347 353 POLY-GLN.  
 FT DOMAIN 357 361 POLY-GLN.  
 FT DOMAIN 410 414 POLY-GLN.  
 FT DOMAIN 426 432 POLY-GLN.  
 FT DOMAIN 445 453 POLY-GLN.  
 FT DOMAIN 456 459 POLY-GLN.  
 FT DOMAIN 466 474 POLY-GLN.  
 FT DOMAIN 497 517 POLY-ALA.  
 FT DOMAIN 524 529 POLY-SER.  
 FT DOMAIN 549 558 POLY-ALA.  
 FT DOMAIN 639 651 POLY-ALA.  
 FT DOMAIN 717 725 POLY-ALA.  
 FT DOMAIN 797 802 POLY-GLN.  
 FT DOMAIN 820 823 POLY-GLN.  
 FT DOMAIN 826 832 POLY-GLN.  
 FT ZN\_FING 874 886 C2H2-TYPE 1.  
 FT ZN\_FING 902 924 C2H2-TYPE 2.  
 FT ZN\_FING 930 953 C2H2-TYPE 3.  
 FT ZN\_FING 969 992 C2H2-TYPE 4.  
 FT CONFLICT 647 647 A -> R (IN REF. 2).  
 SQ SEQUENCE 1028 AA; 110620 MW; D7068BB2BC0F6F77 CRC64;

Query Match 11.9%; Score 92.5; DB 1; Length 1028;  
 Best Local Similarity 26.7%; Pred. No. 2.8;  
 Matches 40; Conservative 14; Mismatches 61; Indels 35; Gaps 6;

QY 3 LLKVAFAAIVVSGSALAGVFWQGGGNGHNGGNSGPDSTLSIYQYGSANAALALQSD 62  
 DB 59 LQNAFAAAIVMAGSG-----GGCTGNGGGGASGPGGGSANSGGGGG----- 104  
 QY 63 ARKSETTITQSGYNGADYVQAGDANSTIELTONGFRFNATIDQWAKYDQIVTRVWTHE 122  
 DB 105 -----GGGYINGCGVG-GPNNS---LDGNLLNFASVSNYNESN-----SKFHNH 147  
 QY 123 MAHA-----NOTASDSVMYRVQVFGNNATA 148  
 DB 148 HHQHNNNNNNNGQTSMGMHFFYGNPSA 177

RESULT 8  
 ID\_TNKL\_HUMAN STANDARD; PRT; 1327 AA.  
 AC Q95271; O95272;  
 DT 28-FEB-2003 (Rel. 41, Created)  
 DT 28-FEB-2003 (Rel. 41, Last sequence update)  
 DT 10-OCT-2003 (Rel. 42, Last annotation update)

DE Tankyrase 1 (EC 2.4.2.30) (TANK1) (Tankyrase I) (TNKS-1) (TRF1-  
 DE interacting ankyrin-related ADP-ribose polymerase).  
 GN TNKS OR TNKS1 OR TINI OR TINI1 OR PARPL.  
 OS Homo sapiens (Human).  
 CC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
 CC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
 CC NCBI\_TaxID=9606;  
 CC [1]  
 RN SEQUENCE FROM N.A. (ISOFORMS 1 AND 2).  
 RN TISSUE=Testis;  
 RC MEDLINE=99040105; PubMed=9822378;  
 RA Smith S., Giriat I., Schmitt A., de Lange T.;  
 RT "Tankyrase, a poly(ADP-ribose) polymerase at human telomeres.";  
 RL Science 282:1484-1487(1998).  
 RN [2]  
 RN SUBCELLULAR LOCATION.  
 RP MEDLINE=99454782; PubMed=10523501;  
 RA Smith S., de Lange T.;  
 RT "Cell cycle dependent localization of the telomeric PARP, tankyrase,  
 RT to nuclear pore complexes and centrosomes.";  
 RL J. Cell Sci. 112:3649-3656(1999).  
 RN [3]  
 RN FUNCTION, AND PHOSPHORYLATION.  
 RP MEDLINE=20556282; PubMed=10988299;  
 RA Chi N.-W., Lodish H.F.;  
 RT "Tankyrase is a Golgi-associated mitogen-activated protein kinase  
 RT substrate that interacts with IRAP in GLUT4 vesicles.";  
 RL J. Biol. Chem. 275:38437-38444(2000).  
 RN [4]  
 RN FUNCTION, AND MUTAGENESIS OF HIS-1184 AND GLU-1291.  
 RP MEDLINE=21602874; PubMed=1179745;  
 RA Cook B.D., Dynek J.N., Chang W., Shostak G., Smith S.;  
 RT "Role for the related poly(ADP-Ribose) polymerases tankyrase 1 and 2  
 RT at human telomeres.";  
 RL Mol. Cell. Biol. 22:332-342(2002).  
 CC -1- FUNCTION: May regulate vesicle trafficking and modulate the  
 CC subcellular distribution of SLC2A4/GLUT4-vesicles. Has PARP  
 CC activity and can modify TRF1, and thereby contribute to the  
 CC regulation of telomere length.  
 CC -1- CATALYTIC ACTIVITY: NAD(+) + {ADP-D-ribose} (N)-acceptor =  
 CC nicotinamide + {ADP-D-ribose} (N+1)-acceptor.  
 CC -1- SUBUNIT: Oligomerizes and associates with TNKS2. Interacts with  
 CC the cytoplasmic domain of LAMP/OTase in SLC2A4/GLUT4-vesicles.  
 CC Binds to the N-terminus of telomeric TRF1 via the ANK repeats.  
 CC -1- SUBCELLULAR LOCATION: Cytoplasmic, associated with the Golgi and  
 CC with juxtanuclear SLC2A4/GLUT4-vesicles. A minor proportion is  
 CC also found at nuclear pore complexes and around the pericentriolar  
 CC matrix of mitotic centrosomes. During interphase, a small fraction  
 CC of TNKS is found in the nucleus, associated with TRF1.  
 CC -1- ALTERNATIVE PRODUCTS:  
 CC Event=Alternative splicing; Named isoforms=2;  
 CC Name=1;  
 CC IsoId=O95271-1; Sequence=Displayed;  
 CC Name=2;  
 CC IsoId=O95271-2; Sequence=VSP\_004538, VSP\_004539;  
 CC Note=No experimental confirmation available;  
 CC -1- TISSUE SPECIFICITY: Ubiquitous; highest levels in testis.  
 CC -1- PTM: Upon insulin-stimulation, phosphorylated on serine residues  
 CC by MAPK kinases.  
 CC -1- PTM: ADP-ribosylated (-auto).  
 CC -1- SIMILARITY: Belongs to the PARP family.  
 CC -1- SIMILARITY: Contains 15 ANK repeats.  
 CC -1- SIMILARITY: Contains 1 sterile alpha motif (SAM) domain.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 CC EMBL; AF082556; AAC79841.1; -

[illegible]

1000-0000  
ID\_VG38\_BPT2  
AC P07875;  
STANDARD; PRT; 262 AA.

DT 01-AUG-1988 (Rel. 08, Created)  
 DT 01-AUG-1988 (Rel. 08, Last sequence update)  
 DT 01-AUG-1990 (Rel. 15, Last annotation update)  
 DE Receptor recognizing protein (Protein Gp38).  
 GN 38.  
 OS Bacteriophage T2.  
 OC Viruses; dsDNA viruses, no RNA stage; Caudovirales; Myoviridae;  
 OC T4-like viruses.  
 OX NCBI\_TaxID=10664;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=87283911; PubMed=3302276;  
 RA Riede I., Drexler K., Eschbach M.L., Henning U.;  
 RT "DNA sequence of genes 38 encoding a receptor-recognizing protein of  
 bacteriophages T2, K3 and of K3 host range mutants.";  
 RL J. Mol. Biol. 194;31-39(1987).  
 CC -!- FUNCTION: Vg38 is at the tip of the long tail fibers and serves as  
 the phage recognition site for the cellular receptor.  
 CC -!- MISCELLANEOUS: THIS PHAGE USE OUTER MEMBRANE PROTEINS OMPF AND TTR  
 AS RECEPTORS.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 the European Bioinformatics Institute. There are no restrictions on its  
 use by non-profit institutions as long as its content is in no way  
 modified and this statement is not removed. Usage by and for commercial  
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL; X05312; CAA28935.1; --  
 DR PIR; S00275; S00275.  
 DR InterPro; IPR007932; Tail fibre\_Gp38.  
 DR Pfam; PF05268; GP38; 1.  
 KW Fiber protein; Phage recognition.  
 FT SIGNAL 1 112 129  
 FT PROPEP 40 87  
 FT CHAIN 88 565  
 FT METAL 251 251  
 FT ACT SITE 252 252  
 FT METAL 255 255  
 FT METAL 321 321  
 FT DISULFID 112 129  
 FT DISULFID 178 217  
 FT DISULFID 301 373  
 FT DISULFID 380 443  
 FT DISULFID 393 412  
 FT DISULFID 402 477  
 FT DISULFID 454 498  
 FT DISULFID 503 553  
 FT DISULFID 523 546  
 FT CARBOHYD 287 287  
 FT LIPID 565 565  
 SQ SEQUENCE 262 AA; 25801 MW; 0567366918F6C745 CRC64;  
  
 Query Match 11.5%; Score 89; DB 1; Length 262;  
 Best Local Similarity 34.4%; Pred. No. 1.2;  
 Matches 32; Conservative 9; Mismatches 38; Indels 14; Gaps 5;  
  
 QY 27 GGGGHHNGGNSGDPSTLSIYQGSANAALQSDARKSETITQSGVNGADYGQAD 86  
 DB 175 GGGGRFPFGVGGKIGSDSILS-----GSNASL---IDAGTGGTTF-QYGAGNGGNGVGGG 225  
 QY 87 ----NSTIETQNGFRNNAIDOWNAKNYDQVLT 116  
 DB 226 RGMGNKVTSEGGAGAAVTG--NAPNMQNVGT 256  
  
 RESULT 11  
 ID GP63\_LEIDO STANDARD; PRT; 590 AA.  
 AC P23223;  
 DT 01-NOV-1991 (Rel. 20, Created)  
 DT 01-NOV-1991 (Rel. 20, Last sequence update)  
 DT 10-OCT-2003 (Rel. 42, Last annotation update)  
 DE Leishmanolysin precursor (EC 3.4.24.36) (Cell surface protease)  
 DE (Major surface glycoprotein) (GP63 protein) (Promastigote surface  
 endopeptidase).  
 DE endopeptidase).  
 GN GP63.  
 OS Leishmania donovani.  
 OC Eukaryota; Euglenozoa; Kinetoplastida; Trypanosomatidae; Leishmania.  
 OX NCBI\_TaxID=5661;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=LV9;  
 RX MEDLINE=92107220; PubMed=1762629;  
 RA Webb J.R., Button L.L., McMaster R.W.;  
 RT "Heterogeneity of the genes encoding the major surface glycoprotein  
 of Leishmania donovani.";  
 RL Mol. Biochem. Parasitol. 48:173-184(1991).  
 CC -!- FUNCTION: Has an integral role during the infection of macrophages  
 in the mammalian host.

CC -!- CATALYTIC ACTIVITY: Preference for hydrophobic residues at P1 and  
 P1' and basic residues at P2 and P3'. A model nonapeptide is  
 cleaved at -Ala-Tyr-[Leu-Lys-Lys-  
 CC -!- COPACTOR: Binds 1 zinc ion per subunit (By similarity).  
 CC -!- SUBCELLULAR LOCATION: Attached to the membrane by a GPI-anchor.  
 CC -!- SIMILARITY: Belongs to peptidase family M8.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 the European Bioinformatics Institute. There are no restrictions on its  
 use by non-profit institutions as long as its content is in no way  
 modified and this statement is not removed. Usage by and for commercial  
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL; M60048; AAA29244.1; --  
 DR HSP; P08148; ILMML.  
 DR MEROPS; M08.001;  
 DR InterPro; IPR006025; Pept\_M\_zn\_BS.  
 DR InterPro; IPR001577; Peptidase M8.  
 DR Pfam; PF01457; Peptidase M8; 1.  
 DR PRINTS; PR00782; ISHMANOLYSIN.  
 DR PROSITE; PS00142; ZINC PROTEASE; 1.  
 KW Hydrolase; Metalloprotease; Glycoprotein; Metal-binding; Zinc;  
 Zymogen; Signal; Cell adhesion; GPI-anchor; Lipoprotein.  
 FT SIGNAL 1 39  
 FT PROPEP 40 87  
 FT CHAIN 88 565  
 FT METAL 251 251  
 FT ACT SITE 252 252  
 FT METAL 255 255  
 FT METAL 321 321  
 FT DISULFID 112 129  
 FT DISULFID 178 217  
 FT DISULFID 301 373  
 FT DISULFID 380 443  
 FT DISULFID 393 412  
 FT DISULFID 402 477  
 FT DISULFID 454 498  
 FT DISULFID 503 553  
 FT DISULFID 523 546  
 FT CARBOHYD 287 287  
 FT LIPID 565 565  
 SQ SEQUENCE 590 AA; 62950 MW; 0FB315D29659F58 CRC64;  
  
 Query Match 11.2%; Score 87; DB 1; Length 590;  
 Best Local Similarity 89.5%; Pred. No. 4.4;  
 Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;  
  
 QY 108 AKNYDQLVTRVTHMAHA 126  
 DB 238 ASRYDQLVTRVTHMAHA 256  
  
 RESULT 12  
 ID GP63\_LEICH STANDARD; PRT; 599 AA.  
 AC P15706;  
 DT 01-APR-1990 (Rel. 14, Created)  
 DT 01-APR-1990 (Rel. 14, Last sequence update)  
 DT 10-OCT-2003 (Rel. 42, Last annotation update)  
 DE Leishmanolysin precursor (EC 3.4.24.36) (Cell  
 surface protease)  
 DE (Major surface glycoprotein) (GP63 protein) (Promastigote surface  
 endopeptidase).  
 DE endopeptidase).  
 GN GP63.  
 OS Leishmania chagasi.  
 OC Eukaryota; Euglenozoa; Kinetoplastida; Trypanosomatidae; Leishmania.  
 OX NCBI\_TaxID=44271;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=90205976; PubMed=2320059;

RA Miller R.A., Reed S.G., Parsons M.;  
 RT "Leishmania gp63 molecule implicated in cellular adhesion lacks an  
 RL Arg-Gly-Asp sequence.";  
 RM Mol. Biochem. Parasitol. 39:267-274(1990).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=92112918; PubMed=1370484;  
 RA Ramamoorthy R., Donelson J.E., Paetz K.E., Maybodi M., Roberts S.C.,  
 RA Wilson M.E.;  
 RT "Three distinct RNAs for the surface protease gp63 are differentially  
 RT expressed during development of Leishmania donovani chagasi  
 RT promastigotes to an infectious form.";  
 RL J. Biol. Chem. 267:1888-1895(1992).  
 CC -!- FUNCTION: Has an integral role during the infection of macrophages  
 CC in the mammalian host.  
 CC -!- CATALYTIC ACTIVITY: Preference for hydrophobic residues at P1 and  
 CC P1' and basic residues at P2 and P3'. A model nonapeptide is  
 CC cleaved at -Ala-Tyr-|-Leu-Lys-Lys-.  
 CC -!- COPACTOR: Binds 1 zinc ion per subunit (By similarity).  
 CC -!- SUBCELLULAR LOCATION: Attached to the membrane by a GPI-anchor.  
 CC -!- SIMILARITY: Belongs to peptidase family M8.  
 CC  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC  
 CC EMBL; M28527; AAA29238.1; -;  
 CC EMBL; M28527; AAA29235.1; -;  
 CC PIR; A44951; A44951.  
 CC HSP; P08148; ILMU.  
 CC MEROPS; M08.001; -;  
 CC InterPro; IPR006025; Pept\_M\_Zn\_BS.  
 CC InterPro; IPR001577; Peptidase\_M8.  
 CC Pfam; PF01457; Peptidase\_M8; 1.  
 CC PRINTS; PR00782; LSHMANOLYSIN.  
 CC PROSITE; PS00142; ZINC PROTEASE; 1.  
 KW Hydrolase; Metalloprotease; Glycoprotein; Metal-binding; Zinc;  
 KW Zymogen; Signal; Cell adhesion; GPI-anchor; Lipoprotein.  
 FT SIGNAL 1 39 POTENTIAL.  
 FT PROPEP 40 97 ACTIVATION PEPTIDE.  
 FT CHAIN 98 574 LEISHMANOLYSIN.  
 FT PROPEP 575 599 REMOVED IN NATURE FORM (BY SIMILARITY).  
 FT METAL 261 261 ZINC (CATALYTIC) (BY SIMILARITY).  
 FT ACT\_SITE 262 262 BY SIMILARITY.  
 FT METAL 265 265 ZINC (CATALYTIC) (BY SIMILARITY).  
 FT METAL 331 331 ZINC (CATALYTIC) (BY SIMILARITY).  
 FT DISULFID 122 139 BY SIMILARITY.  
 FT DISULFID 188 227 BY SIMILARITY.  
 FT DISULFID 311 383 BY SIMILARITY.  
 FT DISULFID 390 452 BY SIMILARITY.  
 FT DISULFID 403 422 BY SIMILARITY.  
 FT DISULFID 412 486 BY SIMILARITY.  
 FT DISULFID 463 507 BY SIMILARITY.  
 FT DISULFID 512 562 BY SIMILARITY.  
 FT DISULFID 532 555 BY SIMILARITY.  
 FT CARBOHYD 297 297 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 394 394 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT LIPID 574 574 GPI-anchor amidated asparagine (By  
 FT similarity).  
 SQ SEQUENCE 599 AA; 63848 MW; 746730AE8E2A2E7C CRC64;  
 Query Match 11.2%; Score 87; DB 1; Length 599;  
 Best Local Similarity 89.5%; Pred. No. 4.4;  
 Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;  
 QY 108 AKNDQDLVTRVWTHEWAHA 126  
 Db 248 ASRYDQLVTRVWTHEWAHA 266

RESULT 13  
 ID GP63 LEIMA STANDARD; PRT; 602 AA.  
 AC P08148; P15906;  
 DT 01-AUG-1988 (Rel. 08, Created)  
 DT 10-APR-1990 (Rel. 14, Last sequence update)  
 DT 10-OCT-2003 (Rel. 42, Last annotation update)  
 DE Leishmanolysin precursor (EC 3.4.24.36) (Cell surface protease)  
 DE (Major surface glycoprotein) (GP63 protein) (Promastigote surface  
 DE endopeptidase).  
 GN GP63.  
 OS Leishmania major.  
 OC Eukaryota; Euklenozoa; Kinetoplastida; Trypanosomatidae; Leishmania.  
 OX NCBI\_TaxID=5664;  
 RN [1]  
 RP SEQUENCE FROM N.A., AND SEQUENCE OF 101-123.  
 RX MEDLINE=88154764; PubMed=3346625;  
 RA Button L.L., McMaster W.R.;  
 RT "Molecular cloning of the major surface antigen of leishmania.";  
 RL J. Exp. Med. 167:724-729(1988).  
 RN [2]  
 RP REVISIONS.  
 RA Button L.L., McMaster W.R.;  
 RL J. Exp. Med. 171:589-589(1990).  
 RN [3]  
 RP GPI-ANCHOR.  
 RX MEDLINE=91009116; PubMed=2145267;  
 RA Schneider P., Ferguson M.A.J., McConville M.J., Mehler A.,  
 RA Homans S.W., Bordier C.;  
 RT "Structure of the glycosyl-phosphatidylinositol membrane anchor of  
 RT the Leishmania major promastigote surface protease.";  
 RL J. Biol. Chem. 265:16955-16964(1990).  
 RN [4]  
 RP X-RAY CRYSTALLOGRAPHY (2.6 ANGSTROMS).  
 RX MEDLINE=95406217; PubMed=7675788;  
 RA Schlagenhaut E., Egges R., Metcalfe P.;  
 RT "Crystallization and preliminary X-ray diffraction studies of  
 RT leishmanolysin, the major surface metalloproteinase from Leishmania  
 RT major.";  
 RL Proteins 22:58-66(1995).  
 RN [5]  
 RP X-RAY CRYSTALLOGRAPHY (1.86 ANGSTROMS).  
 RX MEDLINE=98418698; PubMed=9739094;  
 RA Schlagenhaut E., Egges R., Metcalfe P.;  
 RT "The crystal structure of the Leishmania major surface proteinase  
 RT leishmanolysin.";  
 RL Structure 6:1035-1046(1998).  
 CC -!- FUNCTION: Has an integral role during the infection of macrophages  
 CC in the mammalian host.  
 CC -!- CATALYTIC ACTIVITY: Preference for hydrophobic residues at P1 and  
 CC P1' and basic residues at P2 and P3'. A model nonapeptide is  
 CC cleaved at -Ala-Tyr-|-Leu-Lys-Lys-.  
 CC -!- COPACTOR: Binds 1 zinc ion per subunit.  
 CC -!- SUBCELLULAR LOCATION: Attached to the membrane by a GPI-anchor.  
 CC -!- PTM: THE PHOSPHATIDYLINOSITOL MOIETY OF THE GPI-ANCHOR CONTAINS A  
 CC FULLY SATURATED, UNBRANCHED 1-O-ALKYL CHAIN (MAINLY C24:0) AND A  
 CC MIXTURE OF FULLY SATURATED UNBRANCHED 2-O-ACYL CHAINS (C12:0,  
 CC C14:0, C16:0, AND C18:0).  
 CC -!- SIMILARITY: Belongs to peptidase family M8.  
 CC  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC  
 CC EMBL; Y00647; CAB68673.1; -;  
 CC PIR; PLO221; PLO221.  
 CC PDB; 1LMU; 17-SEP-97.  
 CC MEROPS; M08.001; -;  
 CC

FT	TURN	365	366
FT	STRAND	369	370
FT	HELIX	372	374
FT	TURN	380	383
FT	HELIX	386	390
FT	STRAND	394	395
FT	TURN	396	397
FT	STRAND	398	399
FT	TURN	402	404
FT	STRAND	413	414
FT	TURN	417	418
FT	STRAND	421	425
FT	STRAND	428	429
FT	HELIX	435	437
FT	TURN	443	444
FT	STRAND	445	446
FT	TURN	450	454
FT	STRAND	458	465
FT	TURN	466	467
FT	HELIX	470	472
FT	TURN	475	477
FT	HELIX	478	480
FT	TURN	485	486
FT	STRAND	487	494
FT	STRAND	496	496
FT	STRAND	506	516
FT	TURN	517	520
FT	STRAND	521	525
FT	TURN	527	528
FT	STRAND	533	534
FT	TURN	537	538
FT	STRAND	540	542
FT	HELIX	543	545
FT	TURN	546	546
FT	STRAND	550	550
FT	TURN	552	553
FT	STRAND	555	557
FT	HELIX	561	565
FT	TURN	566	567
FT	HELIX	569	572
FT	TURN	573	573
SS	SEQUENCE	602 AA;	63953 MW; 982EF3245D87C43E CRC64;
Query Match 11.2%; Score 87; DB 1; Length 602;			
Best Local Similarity 89.5%; Prod. No. 4.5;			
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps			
QY	108 AKNYDQLVTRVVTHEMAHA 126		
Db	251 ASRYDQLVTRVVTHEMAHA 269		
RESULT 14			
GP63 LEIME	STANDARD; PRT; 646 AA.		
ID	GP63 LEIME		
AC	P43150;		
DT	01-NOV-1995 (Rel. 32, Created)		
DD	01-NOV-1995 (Rel. 32, Last sequence update)		
DT	10-OCT-2003 (Rel. 42, Last annotation update)		
DE	Leishmanolysin c1 precursor (EC 3.4.24.36) (Cell surface protease)		
DE	(Major surface glycoprotein) (GP63 protein) (Promastigote surface de		
DE	endopeptidase).		
GN	GP63-CL.		
OS	Leishmania mexicana.		
OC	Eukaryota; Euglenozoa; Kinetoplastida; Trypanosomatidae; Leishmania.		
OX	NCBI TaxID=5665;		
RP	[1]		
RP	SEQUENCE FROM N.A.		
RC	STRAIN=MNYC/BZ/62/M379;		
RC	MEDLINE=93149206; PubMed=8426614;		
RR	Medina-Acosta E., Kares R.E., Russell D.G.;		
RT	"Structurally distinct genes for the surface protease of Leishmania		
RT	mexicana are developmentally regulated.";		

Search completed: August 2, 2004, 14:49:31  
Job time : 5.3 secs

This page Blank (uspto)



GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: August 2, 2004, 14:39:17 ; Search time 29.7 seconds  
(without alignments)  
1604.150 Million cell updates/sec

Title: US-09-543-407-24  
Perfect score: 775  
Sequence: 1 MKLLKVAFAAIVVSGSALA.....DSSVMVRQVGFNNATANQY 151

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 1017041 seqs, 315518202 residues

Total number of hits satisfying chosen parameters: 1017041

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

- Database : SPTREMBL\_25.\*
- 1: sp\_archaea.\*
  - 2: sp\_bacteria.\*
  - 3: sp\_fungi.\*
  - 4: sp\_human.\*
  - 5: sp\_invertebrate.\*
  - 6: sp\_mammal.\*
  - 7: sp\_mhc.\*
  - 8: sp\_organelle.\*
  - 9: sp\_phase.\*
  - 10: sp\_plant.\*
  - 11: sp\_rodent.\*
  - 12: sp\_virus.\*
  - 13: sp\_vertebrate.\*
  - 14: sp\_unclassified.\*
  - 15: sp\_rvirus.\*
  - 16: sp\_bacteriaph.\*
  - 17: sp\_archaeap.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	681	87.9	152	2	O33802
2	591.5	76.3	150	2	Q7X243
3	553	71.4	149	2	Q7X240
4	508.5	65.6	152	16	Q8CW63
5	431.5	55.7	150	2	Q7X237
6	385	49.7	76	2	Q54069
7	122	15.7	29	2	Q9S375
8	122	15.7	502	16	Q8E1H4
9	109	14.1	139	16	Q8E1H3
10	104	13.4	7716	16	Q7UWZ8
11	103	13.3	362	16	Q89D03
12	101	13.0	409	5	Q19414
13	100	12.9	151	2	Q7X238
14	100	12.9	362	16	Q8E1H4
15	99.5	12.8	171	16	Q89D13
16	99	12.8	1748	5	Q94821

17	99	12.6	713	5	Q9N8N5	Q9N8N5 trypanosoma
18	96.5	12.5	151	16	Q7UCZ1	Q7UCZ1 shigella fl
19	96.5	12.5	160	16	Q83RU7	Q83RU7 shigella fl
20	96	12.4	154	16	Q89J15	Q89J15 bradyrhizob
21	96	12.4	157	16	Q88HG0	Q88HG0 pseudomonas
22	95.5	12.3	151	2	Q7X244	Q7X244 citrobacter
23	95.5	12.3	180	16	Q8CW64	Q8CW64 escherichia
24	95.5	12.3	624	3	Q8N1V1	Q8N1V1 neurospora
25	95.5	12.3	1286	2	Q841Y5	Q841Y5 campylobact
26	94.5	12.2	453	5	Q9NGM8	Q9NGM8 drosophila
27	94	12.1	179	2	Q33801	Q33801 salmonella
28	93.5	12.1	453	5	Q9NGF6	Q9NGF6 drosophila
29	93.5	12.1	453	5	Q9NGF7	Q9NGF7 drosophila
30	93.5	12.1	552	16	Q8E840	Q8E840 mycobacteri
31	93.5	12.1	552	16	Q71W76	Q71W76 mycobacteri
32	93.5	12.1	623	16	Q8V1Y0	Q8V1Y0 mycobacteri
33	93	12.0	3501	16	Q8Y1O6	Q8Y1O6 raietonia s
34	93	12.0	3552	16	Q8XSD6	Q8XSD6 raietonia s
35	92.5	11.9	348	13	Q933P7	Q933P7 cyprinus ca
36	92.5	11.9	440	16	Q92DG1	Q92DG1 listeria in
37	92.5	11.9	1222	5	Q9W4F0	Q9W4F0 drosophila
38	92.5	11.9	1222	5	Q8T8L9	Q8T8L9 drosophila
39	92.5	11.9	1351	5	Q8SX56	Q8SX56 drosophila
40	92.5	11.9	1354	5	Q8MPN4	Q8MPN4 drosophila
41	91.5	11.8	1422	16	Q8EFU3	Q8EFU3 shewanella
42	91.5	11.8	1615	2	Q9KX28	Q9KX28 rickettsia
43	91	11.7	191	3	Q8TFA6	Q8TFA6 saccharomyc
44	90.5	11.7	152	2	Q7X241	Q7X241 citrobacter
45	90.5	11.7	1209	16	Q89CK5	Q89CK5 bradyrhizob

ALIGNMENTS

RESULT 1

O33802 PRELIMINARY; PRT; 152 AA.  
ID O33802  
AC O33802; (TREMREL. 05, Created)  
DT 01-JAN-1998 (TREMREL. 05, Last sequence update)  
DT 01-JAN-1998 (TREMREL. 05, Last sequence update)  
DT 01-DEC-2001 (TREMREL. 19, Last annotation update)  
DE AgfA protein (Fragment).  
GN AGFA.  
OS Salmonella typhimurium.  
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;  
OC Enterobacteriaceae; Salmonella.  
OX NCBI\_TaxID=602;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=98053981; PubMed=9393832;  
RA Sukupolvi S.S., Lorentz R.G., Gordon J.I., Bian Z., Pfeifer J.D.,  
RA Normark S.J., Rhen M.;  
RT "Expression of thin, aggregative fimbriae promotes interaction of  
RT Salmonella typhimurium SR-11 with mouse small intestinal epithelial  
RT cells.";  
RL Infect. Immun. 65:5320-5325(1997).  
DR EMBL; AJ000514; CAA04151.1; -.  
FT NON\_TER 152  
SQ SEQUENCE 152 AA; 15401 MW; 9DA7DADC2364B006 CRC64;

Query Match 87.9%; Score 681; DB 2; Length 152;  
Best Local Similarity 90.1%; Pred. No. 2e-46; 12; Indels 0; Gaps 0;  
Matches 136; Conservative 3; Mismatches 12;

QY	1	MKLLKVAFAAIVVSGSALA	GVVPPQWGGGNGHNGGNSGDPSTLSITVQYGSANAALAIQ	60
DB	1	MKLLKVAFAAIVVSGSALA	GVVPPQWGGGNGHNGGNSGDPSTLSITVQYGSANAALAIQ	60
QY	61	SDARKSETTITGSGVNGADVCGGADN	STIELTONGFRNNATIDOWNAKNYDQLVTRVVT	120
DB	61	SDARKSETTITGSGVNGADVCGGADN	STIELTONGFRNNATIDOWNAKNSDITVQYGG	120
QY	121	HEMALFANQATSDSSVMVRQVGFNNATANQY	151	



RT "Production of Cellulose and Curli Fimbriae by Members of the Family  
 RT Enterobacteriaceae isolated from the Human Gastrointestinal Tract."  
 RL Infect. Immun. 72:4151-4158 (2003).  
 DR EMBL; AJ515702; CAD56678.1; -  
 SQ SEQUENCE 150 AA; 15112 MW; 508B2D872DF15F3 CRC64;  
 Query Match 55.7%; Score 431.5; DB 2; Length 150;  
 Best Local Similarity 59.6%; Pred. No. 1e-26;  
 Matches 90; Conservative 27; Mismatches 33; Indels 1; Gaps 1;  
 QY 1 MKLLKVAFAIAIVVSGSLAGVDPQWGGGNGHNGGNSGPDSTLSIYQYGSANAALQ 60  
 DB 1 MKFIKVAALAAIVVSGSAGAMINQ-GGWGHGHGGYGGFNSTLNIYQYGGNSALQ 59  
 QY 61 SDARKSETTITQSGYNGADVGQGDNSTIELTQNGFRNNATIDQNAKNYDQLVTRVVT 120  
 DB 60 TDARNSVLNISQTGGGADVGQGDSDSSINLTQNGFGNSATLDQWNSKDSVNVVQYGG 119  
 QY 121 HEMAHANOTASDSSVMVROVCFGNATANQY 151  
 DB 120 LNALVDQTSNSTVNTQIGFGNHATAHQY 150  
 RESULT 6  
 Q54069  
 ID Q54069 PRELIMINARY; PRT; 76 AA.  
 AC Q54069  
 DT 01-NOV-1996 (TRENBLrel. 01, Created)  
 DT 01-NOV-1996 (TRENBLrel. 01, Last sequence update)  
 DT 01-DEC-2001 (TRENBLrel. 19, Last annotation update)  
 DE SEF17 fimbria (Fragment).  
 GN AGFA.  
 OS Salmonella enteritidis.  
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;  
 OC Enterobacteriaceae; Salmonella.  
 OX NCBI\_TaxID=592;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=SE30;  
 RA Cox J.M., Eglezos S., Woolcock J.B.;  
 RT "Virulence of Salmonella enteritidis in chickens correlates with  
 RT colony morphology and expression of SEF17 fimbriae."  
 RL Submitted (Apr-1996) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; U53207; AAA98671.1; -  
 FT NON TER 1  
 FT NON TER 76  
 SQ SEQUENCE 76 AA; 7704 MW; 2FD5411241A7BCB1 CRC64;  
 Query Match 49.7%; Score 385; DB 2; Length 76;  
 Best Local Similarity 98.7%; Pred. No. 2.1e-23;  
 Matches 75; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
 QY 30 GNHGGGNSGPDSTLSIYQYGSANAALQSDARKSETTITQSGYNGADVGQGDNST 89  
 DB 1 GNHGGGNSGPDSTLSIYQYGSANAALQSDARKSETTITQSGYNGADVGQGDNST 60  
 QY 90 IELTQNGFRNNATIDQ 105  
 DB 61 IELTQNGFRNNATIDQ 76  
 RESULT 7  
 Q9S3J5  
 ID Q9S3J5 PRELIMINARY; PRT; 29 AA.  
 AC Q9S3J5  
 DT 01-MAY-2000 (TRENBLrel. 13, Created)  
 DT 01-MAY-2000 (TRENBLrel. 13, Last sequence update)  
 DT 01-MAR-2001 (TRENBLrel. 16, Last annotation update)  
 DE Curlin subunit monomer (Fragment).  
 GN CSGA.  
 OS Escherichia coli.  
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;  
 OC Enterobacteriaceae; Escherichia.

OX NCBI\_TaxID=562;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TRANSPOSON-Insertion sequence IS1;  
 RX MEDLINE=99314153; PubMed=10386375;  
 RA La Ragione R.M., Collighan R.J., Woodward M.J.;  
 RT "Non-curlation of Escherichia coli O78:K80 isolates associated with  
 RT IS1 inserti on in csgB and reduced persistence in poultry infection."  
 RL FEMS Microbiol. Lett. 175:247-253(1999).  
 DR EMBL; AJ131756; CAB45380.1; -  
 FT NON TER 29  
 FT NON TER 29  
 SQ SEQUENCE 29 AA; 2789 MW; E290DFC07ABBE243 CRC64;  
 Query Match 15.7%; Score 122; DB 2; Length 29;  
 Best Local Similarity 89.7%; Pred. No. 0.0042;  
 Matches 26; Conservative 1; Mismatches 2; Indels 0; Gaps 0;  
 QY 1 MKLLKVAFAIAIVVSGSALAGVDPQWGGG 29  
 DB 1 MKLLKVAIAIAIVPSGSALAGVDPQYGGG 29  
 RESULT 8  
 Q8EIH4  
 ID Q8EIH4 PRELIMINARY; PRT; 502 AA.  
 AC Q8EIH4  
 DT 01-MAR-2003 (TRENBLrel. 23, Created)  
 DT 01-MAR-2003 (TRENBLrel. 23, Last sequence update)  
 DT 01-MAR-2003 (TRENBLrel. 23, Last annotation update)  
 DE Conserved hypothetical protein.  
 GN SO0865.  
 OS Shewanella oneidensis.  
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Alteromonadales;  
 OC Alteromonadales; Shewanella.  
 OX NCBI\_TaxID=70863;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=MR-1;  
 RX MEDLINE=22297686; PubMed=12368813;  
 RA Heidelberg J.F., Paulsen I.T., Nelson K.E., Gaidos E.J., Nelson W.C.,  
 RA Read T.D., Eisen J.A., Seshadri R., Ward N., Methe B., Clayton R.A.,  
 RA Meyer T., Tsapin A., Scott J., Beanan M., Brinkac L., Daugherty S.,  
 RA Deboy R.T., Dodson R.J., Durkin A.S., Haft D.H., Kolonay J.F.,  
 RA Madupu R., Peterson J.D., Umayam L.A., White O., Wolf A.M.,  
 RA Vamathevan J., Weidman J., Imprim M., Lee K., Berry K., Lee C.,  
 RA Mueller J., Khouri H., Gill J., Utterback T.R., McDonald L.A.,  
 RA Feldblyum T.V., Smith H.O., Venter J.C., Nealon K.H., Fraser C.M.;  
 RT "Genome sequence of the dissimilatory metal ion-reducing bacterium  
 RT Shewanella oneidensis".  
 RL Nat. Biotechnol. 20:1119-1123(2002).  
 DR EMBL; AE015532; AAN53941.1; -  
 DR TIGR; SO0865; -  
 KW Hypothetical protein; Complete proteome.  
 SQ SEQUENCE 502 AA; 52441 MW; D08CA23D6C46B62D CRC64;  
 Query Match 15.7%; Score 122; DB 16; Length 502;  
 Best Local Similarity 27.0%; Pred. No. 0.12; Indels 32; Gaps 5;  
 Matches 41; Conservative 22; Mismatches 57; Indels 32; Gaps 5;  
 QY 29 GGNHNG-----GGN-----SSGPDSTLSIYQYGSANA---ALALQS 61  
 DB 231 GDNHTGFYALAGSENDISMEQGSNNATYLSMTTGDDNTVDITQGDSTNVGDSLIADI 290  
 QY 62 DARKSETTITQSGYNGADVGQGDNSTIELTQNGFRNNATIDQNAKNYDQLVTRVVT 121  
 DB 291 QGDNDITIKQGDNSGAEFQWVGDSNDVLDKRGDANFATFGAYGTDDFDLSSKGDNN 350  
 QY 122 EMAHANOTASDSSVMVROVCFGN-----NATAN 149  
 DB 351 ELV-AFATGEDNSTEISQEGDANFAYVDATGN 381  
 RESULT 9

Q8EIH3  
ID Q8EIH3 PRELIMINARY; PRT; 139 AA.  
AC Q8EIH3  
DT 01-MAR-2003 (TrEMBLrel. 23, Created)  
DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)  
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)  
DE Minor curilin subunit Ceg8, putative.  
GN S00866.  
OS Shewanella oneidensis.  
OC Bacteria; Proteobacteria; Gammaproteobacteria; Alteromonadales;  
OC Alteromonadaceae; Shewanella.  
OX NCBI\_TaxID=70863;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=MR-1;  
RX MEDLINE=2297686; PubMed=12368813;  
RA Heidelberg J.F., Paulsen I.T., Nelson K.E., Gaidos E.J., Nelson W.C.,  
RA Read T.D., Eissen J.A., Sehadri R., Ward N., Methe B., Clayton R.A.,  
RA Meyer T., Tsapin A., Scott J., Beanan M., Brinkac L., Daugherty S.,  
RA DeSoy R.A., Dodson R.J., Durkin A.S., Haft D.H., Kolonay J.F.,  
RA Madupu R., Peterson J.D., Umayam L.A., White O., Wolf A.M.,  
RA Vamathevan J., Weidman J., Impraim M., Lee K., Berry K., Lee C.,  
RA Mueller H., Khouri H., Gill J., Utterback T.R., McDonald L.A.,  
RA Feildlyum T.V., Smith H.O., Venter J.C., Nealon K.H., Fraser C.M.,  
RA "Genome sequence of the dissimilatory metal ion-reducing bacterium  
RT Shewanella oneidensis";  
RL Nat. Biotechnol. 20:1118-1123 (2002).  
DR EMBL; AE015532; AA053942.1; -;  
DR TIGR; S00866; -;  
KW Complete proteome.  
SQ SEQUENCE 139 AA; 14811 MW; 41EC1CFA76957920 CRC64;

Query Match 14.1%; Score 109; DB 16; Length 139;  
Best Local Similarity 30.1%; Pred. No. 0.28;  
Matches 34; Conservative 19; Mismatches 46; Indels 14; Gaps 3;  
QY 39 SGPSTLSIYQVGSANAALQSDARKSETTITQSGYNGADVGGADNSITELTQNGFR 98  
Db 41 SGRDNLIDVQGTANGIVQSGSDNS-AYVQAGNDNISLVITQIGTNEVQLQVGAQ 99  
QY 99 NNAITQWAKYNDQLVTRVTHEMAHANQTASDSSVMVRQVGFNGNATANY 151  
Db 100 NKASITQIGNDLVQL-----NQLGS-GNFSIQIADGAASITQY 139

RESULT 10  
Q7UWZ8 PRELIMINARY; PRT; 7716 AA.  
ID Q7UWZ8  
AC Q7UWZ8  
DT 01-OCT-2003 (TrEMBLrel. 25, Created)  
DT 01-OCT-2003 (TrEMBLrel. 25, Last sequence update)  
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)  
DE Hypothetical protein.  
GN R81661.  
OS Rhodospirillum rubrum.  
OC Bacteria; Planctomycetes; Planctomycetacia; Planctomycetales;  
OC Planctomycetaceae; Pirellula.  
OX NCBI\_TaxID=117;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=1;  
RX MEDLINE=22735913; PubMed=12835416;  
RA Gloeckner F.O., Kube M., Bauer M., Teeling H., Lombardot T.,  
RA Ludwig W., Gade D., Beck A., Borzym K., Heitmann K., Rabus R.,  
RA Schlesner H., Amann R., Reinhardt R.,  
RT "Complete genome sequence of the marine planctomycete Pirellula sp.  
RT strain 1";  
RL Proc. Natl. Acad. Sci. U.S.A. 100:8298-8303 (2003).  
DR EMBL; BX294135; CAD72214.1; -;  
KW Hypothetical protein; Complete proteome.  
SQ SEQUENCE 7716 AA; 797868 MW; D391A25BD96405C0 CRC64;

Query Match 13.4%; Score 104; DB 16; Length 7716;  
Best Local Similarity 30.1%; Pred. No. 0.28;  
Matches 34; Conservative 19; Mismatches 46; Indels 14; Gaps 3;

Best Local Similarity 29.1%; Pred. No. 81;  
Matches 39; Conservative 18; Mismatches 57; Indels 20; Gaps 6;  
QY 28 GGGNHN--GGGNSS-----GPDSTLSIYQVGSANAALQSDARKSETTITQSGY 75  
Db 4048 GGGNHN--GGGNSS-----GPDSTLSIYQVGSANAALQSDARKSETTITQSGY 75  
QY 76 GNGADVGGQ---ADNSTIELTQNGFRNNAITDQWAKYNDQLVTRVTHEMAHANQTASD 132  
Db 4104 GNGSDVGNNTLPSGATIQLEADGSFSDPCTIQNLNDGETATETTTVSDGNGTDT 4163  
QY 133 SSVVRQVGFNGNA 146  
Db 4164 TSVAITIG-GNDA 4176

RESULT 11  
Q89D03 PRELIMINARY; PRT; 362 AA.  
ID Q89D03  
AC Q89D03  
DT 01-JUN-2003 (TrEMBLrel. 24, Created)  
DT 01-JUN-2003 (TrEMBLrel. 24, Last sequence update)  
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)  
DE Hypothetical exported glutamine-rich protein.  
GN BL17642.  
OS Bradyrhizobium japonicum.  
OC Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;  
OC Bradyrhizobiaceae; Bradyrhizobium.  
OX NCBI\_TaxID=375;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=USDA 110;  
RX MEDLINE=22484998; PubMed=12597275;  
RA Kaneko T., Nakamura Y., Sato S., Minamisawa K., Uchiumi T.,  
RA Sasamoto S., Watanabe A., Idesawa K., Iriguchi M., Kawashima K.,  
RA Kohara M., Matsumoto M., Shimo S., Tsuruoka H., Wada T., Yamada M.,  
RA Tabata S.;  
RT "Complete genomic sequence of nitrogen-fixing symbiotic bacterium  
RT Bradyrhizobium japonicum USDA110";  
RL DNA Res. 9:189-197 (2002).  
DR EMBL; AP005962; BAC52907.1; -;  
KW Hypothetical protein; Complete proteome.  
SQ SEQUENCE 362 AA; 39058 MW; D447EE0D4E8433BC CRC64;

Query Match 13.3%; Score 103; DB 16; Length 362;  
Best Local Similarity 31.6%; Pred. No. 2.6;  
Matches 50; Conservative 14; Mismatches 80; Indels 14; Gaps 6;  
QY 2 KLLKVAFAAIVVVGSGALAGVWPQWGGNGHNGSGNPGDST----LSIYQVGSANAAL 57  
Db 3 KLLFLATTAATAVAIATGALAQSSP---STNSNPTTQRPDSTSTPSSSTPSSGSAQTNP 59  
QY 58 ALQSDARKSETTITQSGYNGADVGGADNSITELTQNGFRNNAIT-----DQWAKYNDQ 113  
Db 60 STNSAQTQSPSTGSAAGQTTNGTGT-NTTQAPT-SNNSTNQATQSPSNQTNAPS-DQ 117  
QY 114 LVTRVTHEMAH-ANQTASDSSVMVRQVGFNGNATANQ 150  
Db 118 TQINPTDNPQAQSANPPASGASQAQSPGTGNSNTTAQ 155

RESULT 12  
Q19414 PRELIMINARY; PRT; 409 AA.  
ID Q19414  
AC Q19414  
DT 01-NOV-1996 (TrEMBLrel. 01, Created)  
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)  
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)  
DE F3E9.4 protein.  
GN F3E9.4.  
OS Caenorhabditis elegans.  
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;  
OC Rhabditidae; Peloderinae; Caenorhabditis.

```

OX NCBI_TaxID=6239;
RN [1]
RP SEQUENCE FROM N.A.
RA McMurray A.A.;
RL Submitted (FEB-1996) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=99069613; PubMed=9851916;
RT none;
RA "Genome sequence of the nematode C.elegans: A platform for
RT investigating biology.";
RL Science 282:2012-2018(1998).
DR EMBL; Z69383; CAA93412.1; -.
DR FIR; T20847; T20847.
DR WormPep; F13E9.4; CE05606.
DR InterPro; IPR003677; Onchocerca_Ag.
DR Pfam; PF02520; DUF148; 1
SQ SEQUENCE 409 AA; 43231 MW; B07DF0E4175C5739 CRC64;

Query Match 13.0%; Score 101; DB 5; Length 409;
Best Local Similarity 23.3%; Pred.No.4.3;
Matches 37; Conservative 26; Mismatches 66; Indels 30; Gaps 5;

QY 15 S G S A L A G V V P W G G G N H G G S S G P D S T L S I Y Q Y G S A N A L Q S D A R K S E T T I T Q S G 74
DB 134 S Q Q A S G S M N F G G Q G V G Q N Q F G G S G F S G S G W G S L S S A N G N N Q G -- S S G 191
QY 75 Y G N -- G A D V G G A D N S T I E -- L T Q N G F R N N ----- A T I Q W N A K 109
DB 192 Y Q N N Q R H Q G Q G G H S S S N S V M S N G Y S S G Y G N N G T P S F L N N V S S A A Q D Y I N V 251
QY 110 N Y D Q L V T R V T H E A H A N O T A D S S V M V R V G F G N N A T A 148
DB 252 N K S L T T N Q I N E Q -- A S N W A S N S V Q A Q I Q I Y E T R S A 287

RESULT 13
Q7X238 PRELIMINARY; PRT; 151 AA.
AC Q7X238;
DT 01-OCT-2003 (TrEMBLrel. 25, Created)
DT 01-OCT-2003 (TrEMBLrel. 25, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Nucleation component of curlin monomers.
OS CSGB.
GN Enterobacter sakazakii.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Enterobacter.
OC NCBI_TaxID=28141;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=Fec39;
RA Zogaj X., Bokranz W., Nimitz M., Romling U.;
RT "Production of Cellulose and Curli Fimbriae by Members of the Family
RT Enterobacteriaceae Isolated from the Human Gastrointestinal Tract.";
RL Infect. Immun. 72:4151-4158(2003).
DR EMBL; A4515702; CAD56677.1; -.
SQ SEQUENCE 151 AA; 15985 MW; F0B82BD2A27882B7 CRC64;

Query Match 12.9%; Score 100; DB 2; Length 151;
Best Local Similarity 31.8%; Pred.No.1.6;
Matches 35; Conservative 12; Mismatches 45; Indels 18; Gaps 4;

QY 10 A A T V S G S A L A G V V P W G G G N H G G S S G P D S T L S I Y Q Y G S A N A L Q S D A R K S E T T 69
DB 58 A Q I R Q E G S K L L S V S Q ----- D A G N R A R V D ----- Q S G T Y N I A W I D Q S - G N G N D A G 103
QY 70 I T O S G Y G N G A D V G Q G G A D N S T I E L T Q N G F R N N A T I D Q W N A K Y D Q L V T R V 119
DB 104 I T O D G Y G N S A K I T Q G S G N R A N I T Q Y G T K T A V V V Q ----- K Q S O A I R V I 149

```



GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: August 2, 2004, 14:35:42 ; Search time 44.9 Seconds  
(without alignments)

950.215 Million cell updates/sec

Title: US-09-543-407-26

Perfect score: 782

Sequence: 1 MKLLKVAFAAIVVSGSALA.....DSSVMRVQVGFNNATANYQ 151

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1586107 seqs, 282547505 residues

Total number of hits satisfying chosen parameters: 1586107

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : A\_Geneseq\_29Jan04:\*

1: Geneseq1980s:\*

2: Geneseq1990s:\*

3: Geneseq2000s:\*

4: Geneseq2001s:\*

5: Geneseq2002s:\*

6: Geneseq2003as:\*

7: Geneseq2003bs:\*

8: Geneseq2004s:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	782	100.0	151	3 AAB36353	Aab36353 Agfa::PT3
2	709	90.7	151	3 AAB36350	Aab36350 Agfa::PT3
3	692	88.5	151	2 AAR74625	Aar74625 Agfa sequ
4	682	88.5	151	3 AAB36341	Aab36341 Salmonell
5	687	87.9	151	2 AAW23570	Aaw23570 Salmonell
6	675	86.3	151	3 AAB36349	Aab36349 Agfa::PT3
7	617	78.9	151	3 AAB36354	Aab36354 Agfa::PT3
8	614	78.5	151	3 AAB36346	Aab36346 Agfa::PT3
9	612	78.3	151	3 AAB36347	Aab36347 Agfa::PT3
10	609	77.9	151	3 AAB36352	Aab36352 Agfa::PT3
11	601	76.9	151	3 AAB36351	Aab36351 Agfa::PT3
12	600	76.7	151	3 AAB36355	Aab36355 Agfa::PT3
13	577	73.8	151	3 AAB36348	Aab36348 Agfa::PT3
14	523	66.9	151	3 AAB36343	Aab36343 Escherich
15	518	66.2	151	7 ABR82651	Abr82651 E. coli C
16	507	64.8	120	2 AAR62761	Aar62761 Agfa sequ
17	507	64.8	120	2 AAW23569	Aaw23569 Salmonell
18	445	56.9	142	2 AAR22664	Aar22664 Fibronect
19	373	47.7	122	2 AAR22663	Aar22663 FMB curli
20	188	24.0	45	3 AAB36346	Aab36346 Salmonell
21	132	16.9	22	3 AAB36318	Aab36318 Salmonell
22	123	15.7	23	3 AAB36331	Aab36331 Salmonell
23	123	15.7	23	3 AAB36326	Aab36326 Salmonell
24	123	15.7	23	3 AAB36338	Aab36338 Salmonell
25	113	14.5	24	7 ABR82644	Abr82644 E. coli C

26	112	14.3	151	3 AAB36344	Aab36344 Escherich
27	111	14.2	22	3 AAB36322	Aab36322 Salmonell
28	111	14.2	22	3 AAB36327	Aab36327 Salmonell
29	111	14.2	22	3 AAB36337	Aab36337 Salmonell
30	107	13.7	151	3 AAB36342	Aab36342 Salmonell
31	105	13.4	597	4 AAU8231	Aau8231 Polypepti
32	102	13.0	26	7 ABR82649	Abr82649 E. coli V
33	100	12.8	943	2 AAW64378	Aaw64378 Mycobacte
34	100	12.8	943	2 AAW81745	Aaw81745 M. tuberc
35	100	12.8	943	2 AAY39032	Aay39032 M. tuberc
36	100	12.8	943	2 AAY39175	Aay39175 M. tuberc
37	100	12.8	3300	6 ABUS6445	Abu36445 Protein e
38	98.5	12.6	738	2 AAW56163	Aaw56163 New DNA s
39	97.5	12.5	850	4 ABB65764	Abb65764 Drosophil
40	97.5	12.5	1028	4 ABB62708	Abb62708 Drosophil
41	96.5	12.3	447	3 AAG29728	Aag29728 Arabidops
42	96.5	12.3	468	3 AAG29727	Aag29727 Arabidops
43	96	12.3	19	3 AAB36323	Aab36323 Salmonell
44	96	12.3	19	3 AAB36336	Aab36336 Salmonell
45	96	12.3	19	3 AAB36328	Aab36328 Salmonell

#### ALIGNMENTS

##### RESULT 1

AAB36353

ID AAB36353 standard; protein; 151 AA.

XX AAB36353;

XX 26-FEB-2001 (first entry)

DE Agfa::PT3#8 amino acid sequence SEQ ID NO:26.

XX Salmonella; agfa; chromosomal gene replacement; fimbrin; epitope;

KW vaccine; immune response; immunogen.

XX Salmonella enteritidis.

OS Escherichia coli.

OS Synthetic.

XX WO200060102-A2.

PN 12-OCT-2000.

XX 05-APR-2000; 2000WO-CA000356.

XX 05-APR-1999; 99US-0127888P.

XX (UYVI-) UNIV VICTORIA.

PI White AP, Doran JL, Collison SK, Kay WW;

XX WPI; 2000-672631/65.

XX N-PSDB; AAC64629.

Recombinant agfa gene having a segment replaced by a foreign DNA sequence

which encodes foreign epitope or antigen, expresses recombinant Agfa

protein useful for eliciting immune response in animal.

Dislosure; Page 138; 139pp; English.

The present invention describes a recombinant agfa gene (I) where a

segment of the gene has been replaced by a segment of a foreign DNA

sequence which encodes a foreign epitope or antigen. Also described are:

(1) use of thin aggregative fimbriae (SEF17/TAF) nucleation depended

assembly system of strains of Salmonella, Escherichia coli and

Enterobacteriaceae for the production of fimbriae comprising recombinant

Agfa, CsgA and Agfa-homologue fimbrial subunits, respectively; (2)

directing recombination of a recombinant gene into the chromosome of the

homologous species; (3) directing recombination of a recombinant gene

back into the chromosome of the homologous species, replacing the native

CC copy of that gene; and (4) eliciting an immune response in an animal,  
 CC comprising separating an amino acid polymer comprising a recombinant Agfa  
 CC protein containing a replacement segment or segments of foreign amino  
 CC acid sequence or sequences grown on a Salmonella, E. coli or  
 CC Enterobacteriaceae host cell, from the host cell and introducing the  
 CC polymer into the animal in conjunction with a carrier or diluent. (1) is  
 CC useful for the expression of recombinant Agfa protein which is useful for  
 CC eliciting an immune response in an animal. In a fimbrial presentation  
 CC system the heterologous antigens are presented in high numbers (up to  
 CC 500,000 copies/cell); the hybrid fimbriae protein possesses both the  
 CC immunogenicity and adhesion properties relevant for an efficient live  
 CC vaccine, the carrier fimbrial subunit proteins are usually strong  
 CC immunogens, which may be important for directing an immune response  
 CC against the inserted epitope, and hybrid fimbriae are easy and  
 CC inexpensive to purify in large amount. The present sequence is given in  
 CC the exemplification of the present invention  
 XX  
 CC Sequence 151 AA;

Query Match 100.0%; Score 782; DB 3; Length 151;  
 Best Local Similarity 100.0%; Pred. No. 2.3e-67;  
 Matches 151; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 MLLKVAAPAAIVVSGSALAGVVPQGGGNGGNGSGPSTLSIYQYGSANAALYDQ 60  
 DB 1 MLLKVAAPAAIVVSGSALAGVVPQGGGNGGNGSGPSTLSIYQYGSANAALYDQ 60  
 QY 61 LVTRVVTHEMAHAGYNGADVGQGDNSTIETQNGFRNNATIDQNAKNSDITVGOYGG 120  
 DB 61 LVTRVVTHEMAHAGYNGADVGQGDNSTIETQNGFRNNATIDQNAKNSDITVGOYGG 120  
 QY 121 NNAALVNOTASDSSVMVRQVGFNNATANQY 151  
 DB 121 NNAALVNOTASDSSVMVRQVGFNNATANQY 151

RESULT 2  
 AAB36350  
 ID AAB36350 standard; protein; 151 AA.  
 XX  
 AC AAB36350;  
 XX  
 XX 26-FEB-2001 (first entry)  
 XX  
 XX Agfa:PT3#5 amino acid sequence SEQ ID NO:20.  
 XX  
 KW Salmonella; agfa; chromosomal gene replacement; fimbria; epitope;  
 KW vaccine; immune response; immunogen.  
 XX  
 OS Salmonella enteritidis.  
 OS Escherichia coli.  
 OS Synthetic.  
 XX  
 PN WO2000060102-A2.  
 XX  
 XX 12-OCT-2000.  
 XX  
 XX 05-APR-2000; 2000WO-CA000356.  
 XX  
 XX 05-APR-1999; 99US-0127888P.  
 XX  
 XX (UYVI-) UNIV VICTORIA.  
 XX  
 XX White AP, Doran JL, Collison SK, Kay WW;  
 XX  
 XX WPI: 2000-672631/65.  
 XX N-PSDB; AAC64626.  
 XX  
 XX Recombinant agfa gene having a segment replaced by a foreign DNA sequence  
 XX which encodes foreign epitope or antigen, expresses recombinant Agfa  
 XX protein useful for eliciting immune response in animal.  
 XX  
 XX Disclosure; Page 137; 139pp; English.

XX The present invention describes a recombinant agfa gene (1) where a  
 CC segment of the gene has been replaced by a segment of a foreign DNA  
 CC sequence which encodes a foreign epitope or antigen. Also described are:  
 CC (1) use of thin aggregative fimbriae (SEF17/TAF) nucleation depended  
 CC assembly system of strains of Salmonella, Escherichia coli and  
 CC Enterobacteriaceae for the production of fimbriae comprising recombinant  
 CC Agfa, CsgA and Agfa-homologue fimbria subunits, respectively; (2)  
 CC directing recombination of a recombinant gene into the chromosome of the  
 CC homologous species; (3) directing recombination of a recombinant gene  
 CC back into the chromosome of the homologous species, replacing the native  
 CC copy of that gene; and (4) eliciting an immune response in an animal,  
 CC comprising separating an amino acid polymer comprising a recombinant Agfa  
 CC protein containing a replacement segment or segments of foreign amino  
 CC acid sequence or sequences grown on a Salmonella, E. coli or  
 CC Enterobacteriaceae host cell, from the host cell and introducing the  
 CC polymer into the animal in conjunction with a carrier or diluent. (1) is  
 CC useful for the expression of recombinant Agfa protein which is useful for  
 CC eliciting an immune response in an animal. In a fimbrial presentation  
 CC system the heterologous antigens are presented in high numbers (up to  
 CC 500,000 copies/cell); the hybrid fimbriae protein possesses both the  
 CC immunogenicity and adhesion properties relevant for an efficient live  
 CC vaccine, the carrier fimbrial subunit proteins are usually strong  
 CC immunogens, which may be important for directing an immune response  
 CC against the inserted epitope, and hybrid fimbriae are easy and  
 CC inexpensive to purify in large amount. The present sequence is given in  
 CC the exemplification of the present invention  
 XX  
 CC Sequence 151 AA;

Query Match 90.7%; Score 709; DB 3; Length 151;  
 Best Local Similarity 89.8%; Pred. No. 2.5e-60;  
 Matches 143; Conservative 0; Mismatches 0; Indels 16; Gaps 2;  
 QY 1 MLLKVAAPAAIVVSGSALAGVVPQGGGNGGNGSGPSTLSIYQYGSANAALYDQ 57  
 DB 1 MLLKVAAPAAIVVSGSALAGVVPQGGGNGGNGSGPSTLSIYQYGSANAALYDQ 60  
 QY 58 -----YDQLVTRVVTHEMAHAGYNGADVGQGDNSTIETQNGFRNNATIDQNAKNSD 112  
 DB 61 SDARKYDQLVTRVVTHEMAHA-----CGADNSTIETQNGFRNNATIDQNAKNSD 112  
 QY 113 ITVGQYGGNNAALVNOTASDSSVMVRQVGFNNATANQY 151  
 DB 113 ITVGQYGGNNAALVNOTASDSSVMVRQVGFNNATANQY 151

RESULT 3  
 AAR74625  
 ID AAR74625 standard; protein; 151 AA.  
 XX  
 AC AAR74625;  
 XX  
 XX 25-MAR-2003 (revised)  
 XX 26-JUN-1995 (first entry)  
 XX  
 XX Agfa sequence.  
 XX  
 XX Salmonella; Agfa; vaccine.  
 XX  
 XX Salmonella.  
 XX  
 XX WO9425598-A2.  
 XX  
 XX 10-NOV-1994.  
 XX  
 XX 26-APR-1994; 94WO-IB000207.  
 XX  
 XX 26-APR-1993; 93US-00054452.  
 XX (UYVI-) UNIV VICTORIA INNOVATION & DEV CORP.  
 XX (KING/) KING J.  
 XX



PI Kay WW, Collinson SK, Clouthier SC, Doran JL;  
 XX WPI; 1994-358275/44.  
 DR N-PSDB; AAQ87467.  
 XX  
 PT Eliciting an immune response to Salmonella - using attenuated Salmonella  
 PT strains, vector constructs, or compsns. contg. fimbrial type proteins.  
 XX  
 PS Disclosure; Fig 7B; 95pp; English.  
 XX  
 CC The Salmonella Agfa protein and DNA are used in vaccine and genetic  
 CC immunization compositions, respectively, to elicit an immune response to  
 CC Salmonella in animals (e.g. food producing animals) and humans. (Updated  
 CC on 25-MAR-2003 to correct PN field.)  
 XX  
 SQ Sequence 151 AA;  
 Query Match 88.5%; Score 692; DB 2; Length 151;  
 Best Local Similarity 90.7%; Pred. No. 1.1e-58;  
 Matches 137; Conservative 2; Mismatches 12; Indels 0; Gaps 0;  
 QY 1 MKLLKVAFAAIVVSGSALAGVVPQWGGGNGHNGGNSGPDSTLSIYQYGSANAALYDQ 60  
 DB 1 MKLLKVAFAAIVVSGSALAGVVPQWGGGNGHNGGNSGPDSTLSIYQYGSANAALALQ 60  
 QY 61 LVTRVVTHEMAHAGYNGADVGQADNSTIETONGFRNNATIDOWNAKNSDITVGOYGG 120  
 DB 61 SDARKSETTITQSGYNGADVGQADNSTIETONGFRNNATIDOWNAKNSDITVGOYGG 120  
 QY 121 NNAALVNQTSADSSVMVRQVGFNNATANQY 151  
 DB 121 NNAALVNQTSADSSVMVRQVGFNNATANQY 151  
 RESULT 4  
 AAB36341  
 ID AAB36341 standard; protein; 151 AA.  
 AC AAB36341;  
 XX  
 DT 26-FEB-2001 (first entry)  
 XX  
 DE Salmonella enteritidis Agfa amino acid sequence SEQ ID NO:5.  
 XX  
 KW Salmonella; agfa; chromosomal gene replacement; fimbrin; epitope;  
 KW vaccine; immune response; immunogen.  
 XX  
 OS Salmonella enteritidis.  
 XX  
 PN WO200060102-A2.  
 XX  
 PD 12-OCT-2000.  
 XX  
 PF 05-APR-2000; 2000WO-CA000356.  
 XX  
 PR 05-APR-1999; 99US-0127888P.  
 XX  
 PA (UYVI-) UNIV VICTORIA.  
 XX  
 PI White AP, Doran JL, Collinson SK, Kay WW;  
 XX  
 DR WPI; 2000-672631/65.  
 DR N-PSDB; AAC64617.  
 XX  
 PT Recombinant agfa gene having a segment replaced by a foreign DNA sequence  
 PT which encodes foreign epitope or antigen, expresses recombinant Agfa  
 PT protein useful for eliciting immune response in animal.  
 XX  
 PS Disclosure; Page 135; 139pp; English.  
 XX  
 CC The present invention describes a recombinant agfa gene (I) where a  
 CC segment of the gene has been replaced by a segment of a foreign DNA  
 CC sequence which encodes a foreign epitope or antigen. Also described are:

CC (1) use of thin aggregative fimbriae (SEF17/TAF) nucleation depended  
 CC assembly system of strains of Salmonella, Escherichia coli and  
 CC Enterobacteriaceae for the production of fimbriae comprising recombinant  
 CC Agfa, CsgA and Agfa-homologue fimbrin subunits, respectively; (2)  
 CC directing recombination of a recombinant gene into the chromosome of the  
 CC homologous species; (3) directing recombination of a recombinant gene  
 CC back into the chromosome of the homologous species, replacing the native  
 CC copy of that gene; and (4) eliciting an immune response in an animal  
 CC comprising separating an amino acid polymer comprising a recombinant Agfa  
 CC protein containing a replacement segment or segments of foreign amino  
 CC acid sequence or sequences grown on a Salmonella, E. coli or  
 CC Enterobacteriaceae host cell, from the host cell and introducing the  
 CC polymer into the animal in conjunction with a carrier or diluent. (I) is  
 CC useful for the expression of recombinant Agfa protein which is useful for  
 CC eliciting an immune response in an animal. In a fimbrial presentation  
 CC system the heterologous antigens are presented in high numbers (up to  
 CC 500,000 copies/cell), the hybrid fimbrin protein possesses both the  
 CC immunogenicity and adhesion properties relevant for an efficient live  
 CC vaccine, the carrier fimbrial subunit proteins are usually strong  
 CC immunogens, which may be important for directing an immune response  
 CC against the inserted epitope, and hybrid fimbriae are easy and  
 CC inexpensive to purify in large amount. The present sequence is given in  
 CC the exemplification of the present invention  
 XX  
 SQ Sequence 151 AA;

Query Match 88.5%; Score 692; DB 3; Length 151;  
 Best Local Similarity 90.7%; Pred. No. 1.1e-58;  
 Matches 137; Conservative 2; Mismatches 12; Indels 0; Gaps 0;  
 QY 1 MKLLKVAFAAIVVSGSALAGVVPQWGGGNGHNGGNSGPDSTLSIYQYGSANAALYDQ 60  
 DB 1 MKLLKVAFAAIVVSGSALAGVVPQWGGGNGHNGGNSGPDSTLSIYQYGSANAALALQ 60  
 QY 61 LVTRVVTHEMAHAGYNGADVGQADNSTIETONGFRNNATIDOWNAKNSDITVGOYGG 120  
 DB 61 SDARKSETTITQSGYNGADVGQADNSTIETONGFRNNATIDOWNAKNSDITVGOYGG 120  
 QY 121 NNAALVNQTSADSSVMVRQVGFNNATANQY 151  
 DB 121 NNAALVNQTSADSSVMVRQVGFNNATANQY 151

RESULT 5  
 AAW23570  
 ID AAW23570 standard; protein; 151 AA.  
 XX  
 AC AAW23570;  
 XX  
 DT 25-MAR-2003 (revised)  
 DT 29-SEP-1997 (first entry)  
 XX  
 DE Salmonella enteritidis 27655-3b agfa.  
 XX  
 KW Enteropathogenic bacteria; enterobacteria; S.enteritidis; antibody.  
 XX  
 OS Salmonella enteritidis.  
 XX  
 FH Key Location/Qualifiers  
 FT Misc-difference 123  
 FT /note= "Encoded by GCC"  
 XX  
 PN US5635617-A.  
 XX  
 PD 03-JUN-1997.  
 XX  
 PF 26-APR-1994; 94US-00233788.  
 XX  
 PR 26-APR-1993; 93US-00054452.  
 XX  
 PA (UYVI-) UNIV VICTORIA INNOVATION & DEV CORP.  
 XX  
 PI Collinson SK, Kay WW, Doran JL;

XX WPI: 1997-309886/28.  
 DR N-PSDB; AAR74142.  
 XX  
 PT Isolated Salmonella gene agfa - used for diagnosis of Salmonella or  
 PT enteropathogenic bacteria of the Enterobacteriaceae family.  
 XX  
 PS Example 2; Fig 7; 85pp; English.  
 XX  
 CC The present sequence represents agfa encoded by the full agfa gene  
 CC derived from Salmonella enteritidis 27655-3b. The nucleic acid can be  
 CC used to provide diagnostic assays for Salmonella and/or enteropathogenic  
 CC bacteria of the family Enterobacteriaceae. It can also be used to provide  
 CC proteins and antibodies which can be used for assays. The nucleic acid  
 CC sequence can be used to provide probes or primers which can specifically  
 CC hybridize to nucleic acid molecules from greater than 99% of Salmonella  
 CC strains that are pathogenic to warm-blooded animals relative to nucleic  
 CC acid molecules from virtually all other microbial organisms. (Updated on  
 CC 25-MAR-2003 to correct PF field.)  
 XX  
 SQ Sequence 151 AA;  
 Query Match 87.9%; Score 687; DB 2; Length 151;  
 Best Local Similarity 90.1%; Pred. No. 3.3e-58;  
 Matches 136; Conservative 2; Mismatches 13; Indels 0; Gaps 0;  
 QY 1 MLLKVAFAAIVVSGSALAGVVPQWGGGNGHNGSGSPDSTLSIYQYGSANAALYDQ 60  
 DB 1 MLLKVAFAAIVVSGSALAGVVPQWGGGNGHNGSGSPDSTLSIYQYGSANAALYDQ 60  
 QY 61 LVTRVVTHEMAHAGYQGGADVGQGGADNSTIELTQNGFRNNATIDQNAKNSDITVQYGG 120  
 DB 61 SDARKSETTITQSGYGGADVGQGGADNSTIELTQNGFRNNATIDQNAKNSDITVQYGG 120  
 QY 121 NNAALVNQTASDSSVMVRQVGFNNATANQY 151  
 DB 121 NNPALVNQTASDSSVMVRQVGFNNATANQY 151  
 RESULT 6  
 AAB36349  
 ID AAB36349 standard; protein; 151 AA.  
 XX  
 AC AAB36349;  
 XX  
 DT 26-FEB-2001 (first entry)  
 XX  
 DE Agfa::PT3#4 amino acid sequence SEQ ID NO:18.  
 XX  
 KW Salmonella; agfa; chromosomal gene replacement; fimbrin; epitope;  
 KW vaccine; immune response; immunogen.  
 XX  
 OS Salmonella enteritidis.  
 OS Escherichia coli.  
 OS Synthetic.  
 XX  
 PN W0200060102-A2.  
 XX  
 PD 12-OCT-2000.  
 XX  
 PF 05-APR-2000; 2000WO-CA000356.  
 XX  
 PR 05-APR-1999; 99US-0127888P.  
 XX  
 PA (UUVI-) UNIV VICTORIA.  
 XX  
 PI White AP, Doran JL, Collison SK, Kay WW;  
 XX  
 DR WPI: 2000-672631/65.  
 DR N-PSDB; AAC64625.  
 XX  
 PT Recombinant agfa gene having a segment replaced by a foreign DNA sequence  
 PT which encodes foreign epitope or antigen, expresses recombinant Agfa

PT protein useful for eliciting immune response in animal.  
 XX  
 PS Disclosure; Page 136; 139pp; English.  
 XX  
 CC The present invention describes a recombinant agfa gene (1) where a  
 CC segment of the gene has been replaced by a segment of a foreign DNA  
 CC sequence which encodes a foreign epitope or antigen. Also described are:  
 CC (1) use of thin aggregative fimbriae (SEF17/TAf) nucleation depended  
 CC assembly system of strains of Salmonella, Escherichia coli and  
 CC Enterobacteriaceae for the production of fimbriae comprising recombinant  
 CC Agfa, CsaA and Agfa-homologue fimbrin subunits, respectively; (2)  
 CC directing recombination of a recombinant gene into the chromosome of the  
 CC homologous species; (3) directing recombination of a recombinant gene  
 CC back into the chromosome of the homologous species, replacing the native  
 CC copy of that gene; and (4) eliciting an immune response in an animal,  
 CC comprising separating an amino acid polymer comprising a recombinant Agfa  
 CC protein containing a replacement segment or segments of foreign amino  
 CC acid sequence or sequences grown on a Salmonella, E. coli or  
 CC Enterobacteriaceae host cell, from the host cell and introducing the  
 CC polymer into the animal in conjunction with a carrier or diluent. (1) is  
 CC useful for the expression of recombinant Agfa protein which is useful for  
 CC eliciting an immune response in an animal. In a fimbrial presentation  
 CC system the heterologous antigens are presented in high numbers (up to  
 CC 500,000 copies/cell), the hybrid fimbrin protein possesses both the  
 CC immunogenicity and adhesion properties relevant for an efficient live  
 CC vaccine, the carrier fimbrial subunit proteins are usually strong  
 CC immunogens, which may be important for directing an immune response  
 CC against the inserted epitope, and hybrid fimbriae are easy and  
 CC inexpensive to purify in large amount. The present sequence is given in  
 CC the exemplification of the present invention  
 XX  
 SQ Sequence 151 AA;  
 Query Match 86.3%; Score 675; DB 3; Length 151;  
 Best Local Similarity 81.9%; Pred. No. 4.7e-57;  
 Matches 136; Conservative 0; Mismatches 0; Indels 30; Gaps 2;  
 QY 1 MLLKVAFAAIVVSGSALAGVVPQWGGGNGHNGSGSPDSTLSIYQYGSANAALYDQ 60  
 DB 1 MLLKVAFAAIVVSGSALAGVVPQWGGGNGHNGSGSPDSTLSIYQYGSANAALYDQ 45  
 QY 61 LVTRVVTHEMAHA-----GYGNGADVGQGGADNSTIELTQNGFRNNATIDQ 105  
 DB 46 LVTRVVTHEMAHALQSDARKSETTITQSGYGGADVGQGGADNSTIELTQNGFRNNATIDQ 105  
 QY 106 WNAKNSDITVQYGGNNAALVNQTASDSSVMVRQVGFNNATANQY 151  
 DB 106 WNAKNSDITVQYGGNNAALVNQTASDSSVMVRQVGFNNATANQY 151  
 RESULT 7  
 AAB36354  
 ID AAB36354 standard; protein; 151 AA.  
 XX  
 AC AAB36354;  
 XX  
 DT 26-FEB-2001 (first entry)  
 XX  
 DE Agfa::PT3#9 amino acid sequence SEQ ID NO:28.  
 XX  
 KW Salmonella; agfa; chromosomal gene replacement; fimbrin; epitope;  
 KW vaccine; immune response; immunogen.  
 XX  
 OS Salmonella enteritidis.  
 OS Escherichia coli.  
 OS Synthetic.  
 XX  
 PN W0200060102-A2.  
 XX  
 PD 12-OCT-2000.  
 XX  
 PF 05-APR-2000; 2000WO-CA000356.  
 XX

PR 05-APR-1999; 99US-0127888P.  
XX (UYVI-) UNIV VICTORIA.  
PA White AP, Doran JL, Collison SK, Kay WW;  
PI WPI; 2000-672631/65.  
DR N-PSDB; AAC64630.  
XX Recombinant agfa gene having a segment replaced by a foreign DNA sequence  
PT which encodes foreign epitope or antigen, expresses recombinant Agfa  
PT protein useful for eliciting immune response in animal.  
XX Disclosure; Page 138; 139pp; English.  
PS The present invention describes a recombinant agfa gene (I) where a  
XX segment of the gene has been replaced by a segment of a foreign DNA  
CC sequence which encodes a foreign epitope or antigen. Also described are:  
CC (1) use of thin aggregative fimbriae (SEF17/TAf) nucleation depended  
CC assembly system of strains of Salmonella, Escherichia coli and  
CC Enterobacteriaceae for the production of fimbriae comprising recombinant  
CC Agfa, CsgA and Agfa-homologue fimbrial subunits, respectively; (2)  
CC directing recombination of a recombinant gene into the chromosome of the  
CC homologous species; (3) directing recombination of a recombinant gene  
CC back into the chromosome of the homologous species, replacing the native  
CC copy of that gene; and (4) eliciting an immune response in an animal,  
CC comprising separating an amino acid polymer comprising a recombinant Agfa  
CC protein containing a replacement segment or segments of foreign amino  
CC acid sequence or sequences grown on a Salmonella, E. coli or  
CC Enterobacteriaceae host cell, from the host cell and introducing the  
CC polymer into the animal in conjunction with a carrier or diluent. (I) is  
CC useful for the expression of recombinant Agfa protein which is useful for  
CC eliciting an immune response in an animal. In a fimbrial presentation  
CC system the heterologous antigens are presented in high numbers (up to  
CC 500,000 copies/cell), the hybrid fimbrial protein possesses both the  
CC immunogenicity and adhesion properties relevant for an efficient live  
CC vaccine, the carrier fimbrial subunit proteins are usually strong  
CC immunogens, which may be important for directing an immune response  
CC against the inserted epitope, and hybrid fimbriae are easy and  
CC inexpensive to purify in large amount. The present sequence is given in  
CC the exemplification of the present invention  
XX Sequence 151 AA;  
SQ  
Query Match 78.9%; Score 617; DB 3; Length 151;  
Best Local Similarity 73.6%; Pred. No. 1.8e-51;  
Matches 128; Conservative 0; Mismatches 0; Indels 46; Gaps 2;  
QY 1 MKLLKVAAPAAIIVSGSALAGVVPQWGGGNGGNGGNGSGSPDSTLSIYQYGSANAAL---- 57  
DB 1 MKLLKVAAPAAIIVSGSALAGVVPQWGGGNGGNGGNGSGSPDSTLSIYQYGSANAALALQ 60  
QY 58 -----YDQLVTRVVTHEMAHAGYNGADVGQGDADNSTIELTQNGF 97  
DB 61 SDARKSETTITQSGYNGADYDQLVTRVVTHEMAHA-----F 97  
QY 98 RNNATIDQNNKNSDITVGGYGGNNAALVNQTSASDSSVMVRQVFGNNATANQY 151  
DB 98 RNNATIDQNNKNSDITVGGYGGNNAALVNQTSASDSSVMVRQVFGNNATANQY 151  
RESULT 8  
AAB36346  
ID AAB36346 standard; protein; 151 AA.  
XX  
AC AAB36346;  
XX  
XX 26-FEB-2001 (first entry)  
XX  
XX Agfa::PT3#1 amino acid sequence SEQ ID NO:12.  
XX Salmonella; agfa; chromosomal gene replacement; fimbrial; epitope;  
KW vaccine; immune response; immunogen.

XX Salmonella enteritidis.  
OS Escherichia coli.  
OS Synthetic.  
XX WO2000060102-A2.  
XX 12-OCT-2000.  
XX 05-APR-2000; 2000WO-CA000356.  
XX 05-APR-1999; 99US-0127888P.  
XX (UYVI-) UNIV VICTORIA.  
PA White AP, Doran JL, Collison SK, Kay WW;  
PI WPI; 2000-672631/65.  
DR N-PSDB; AAC64622.  
XX Recombinant agfa gene having a segment replaced by a foreign DNA sequence  
PT which encodes foreign epitope or antigen, expresses recombinant Agfa  
PT protein useful for eliciting immune response in animal.  
XX Disclosure; Page 135; 139pp; English.  
PS The present invention describes a recombinant agfa gene (I) where a  
XX segment of the gene has been replaced by a segment of a foreign DNA  
CC sequence which encodes a foreign epitope or antigen. Also described are:  
CC (1) use of thin aggregative fimbriae (SEF17/TAf) nucleation depended  
CC assembly system of strains of Salmonella, Escherichia coli and  
CC Enterobacteriaceae for the production of fimbriae comprising recombinant  
CC Agfa, CsgA and Agfa-homologue fimbrial subunits, respectively; (2)  
CC directing recombination of a recombinant gene into the chromosome of the  
CC homologous species; (3) directing recombination of a recombinant gene  
CC back into the chromosome of the homologous species, replacing the native  
CC copy of that gene; and (4) eliciting an immune response in an animal,  
CC comprising separating an amino acid polymer comprising a recombinant Agfa  
CC protein containing a replacement segment or segments of foreign amino  
CC acid sequence or sequences grown on a Salmonella, E. coli or  
CC Enterobacteriaceae host cell, from the host cell and introducing the  
CC polymer into the animal in conjunction with a carrier or diluent. (I) is  
CC useful for the expression of recombinant Agfa protein which is useful for  
CC eliciting an immune response in an animal. In a fimbrial presentation  
CC system the heterologous antigens are presented in high numbers (up to  
CC 500,000 copies/cell), the hybrid fimbrial protein possesses both the  
CC immunogenicity and adhesion properties relevant for an efficient live  
CC vaccine, the carrier fimbrial subunit proteins are usually strong  
CC immunogens, which may be important for directing an immune response  
CC against the inserted epitope, and hybrid fimbriae are easy and  
CC inexpensive to purify in large amount. The present sequence is given in  
CC the exemplification of the present invention  
XX Sequence 151 AA;  
SQ  
Query Match 78.5%; Score 614; DB 3; Length 151;  
Best Local Similarity 80.8%; Pred. No. 3.5e-51;  
Matches 122; Conservative 6; Mismatches 23; Indels 0; Gaps 0;  
QY 1 MKLLKVAAPAAIIVSGSALAGVVPQWGGGNGGNGGNGSGSPDSTLSIYQYGSANAALYDQ 60  
DB 1 MKLLKVAAPAAIIVSGSALAGVVPQWGGGNGGNGGNGSGSPDSTLSIYQYGSANAALALQ 60  
QY 61 LVTRVVTHEMAHAGYNGADVGQGDADNSTIELTQNGFNNATIDQNNKNSDITVGGYGG 120  
DB 61 SDARKSETTITQSGYNGADVGQGDADNSTIELTQNGFNNATIDQNNKNSDITVGGYGG 120  
QY 121 NNAALVNQTSASDSSVMVRQVFGNNATANQY 151  
DB 121 NNAALVNQTSASDSSVMVRQVFGNNATANQY 151  
RESULT 9



CC the exemplification of the present invention  
 XX  
 SQ Sequence 151 AA;

Query Match 77.9%; Score 609; DB 3; Length 151;  
 Best Local Similarity 82.1%; Pred. No. 1.1e-50;  
 Matches 124; Conservative 4; Mismatches 23; Indels 0; Gaps 0;

QY 1 MKLLKVAFAAIVVSGSALAGVVPQWGGGNGHNGGSSGPDSTLSIYQYGSANAALYDQ 60  
 DB 1 MKLLKVAFAAIVVSGSALAGVVPQWGGGNGHNGGSSGPDSTLSIYQYGSANAALQ 60

QY 61 LVTRVVTHEMAHAGYNGGADVGQADNSTIELTONGFRNNATIDOWNAKNSDITVGYGG 120  
 DB 61 SDARKSETTITQSGYNGGADVGQADNSTIELTONGFRNNATIDOWNAKNSDITVTRVT 120

QY 121 NNAALVNOTASDSSVMVRQVFGNNATANQY 151  
 DB 121 HEMAHANOTASDSSVMVRQVFGNNATANQY 151

RESULT 11  
 AAB36351  
 ID AAB36351 standard; protein; 151 AA.  
 AC AAB36351;  
 XX  
 DT 26-FEB-2001 (first entry)  
 XX  
 DE Agfa::PT3#6 amino acid sequence SEQ ID NO:22.  
 XX  
 KW Salmonella; agfa; chromosomal gene replacement; fimbrin; epitope;  
 XX vaccine; immune response; immunogen.  
 XX  
 OS Salmonella enteritidis.  
 OS Escherichia coli.  
 OS Synthetic.  
 XX  
 PN WO200060102-A2.  
 XX  
 PD 12-OCT-2000.  
 XX  
 PF 05-APR-2000; 2000WO-CA000356.  
 XX  
 PR 05-APR-1999; 99US-0127888P.  
 XX  
 PA (UYVI-) UNIV VICTORIA.  
 XX  
 PI White AP, Doran JL, Collison SK, Kay WW;  
 XX  
 DR WPI; 2000-672631/65.  
 XX  
 DR N-PSDB; AAC64627.  
 XX  
 PT Recombinant agfa gene having a segment replaced by a foreign DNA sequence  
 PT which encodes foreign epitope or antigen, expresses recombinant Agfa  
 PT protein useful for eliciting immune response in animal.  
 XX  
 PS Disclosure; Page 137; 139pp; English.

The present invention describes a recombinant agfa gene (I) where a segment of the gene has been replaced by a segment of a foreign DNA sequence which encodes a foreign epitope or antigen. Also described are: (1) use of thin aggregative fimbriae (SEF17/TAF) nucleation depended assembly system of strains of Salmonella, Escherichia coli and Enterobacteriaceae for the production of fimbriae comprising recombinant Agfa, CsgA and Agfa-homologue fimbrial subunits, respectively; (2) directing recombination of a recombinant gene into the chromosome of the homologous species; (3) directing recombination of a recombinant gene back into the chromosome of the homologous species, replacing the native copy of that gene; and (4) eliciting an immune response in an animal, comprising separating an amino acid polymer comprising a recombinant Agfa protein containing a replacement segment or segments of foreign amino acid sequence or sequences grown on a Salmonella, E. coli or

CC Enterobacteriaceae host cell, from the host cell and introducing the polymer into the animal in conjunction with a carrier or diluent. (f) is useful for the expression of recombinant Agfa protein which is useful for eliciting an immune response in an animal in a fimbrial presentation system the heterologous antigens are presented in high numbers (up to 500,000 copies/cell), the hybrid fimbrin protein possesses both the immunogenicity and adhesion properties relevant for an efficient live vaccine, the carrier fimbrial subunit proteins are usually strong immunogens, which may be important for directing an immune response against the inserted epitope, and hybrid fimbriae are easy and inexpensive to purify in large amount. The present sequence is given in the exemplification of the present invention

XX  
 SQ Sequence 151 AA;

Query Match 76.9%; Score 601; DB 3; Length 151;  
 Best Local Similarity 81.5%; Pred. No. 6.3e-50;  
 Matches 123; Conservative 4; Mismatches 24; Indels 0; Gaps 0;

QY 1 MKLLKVAFAAIVVSGSALAGVVPQWGGGNGHNGGSSGPDSTLSIYQYGSANAALYDQ 60  
 DB 1 MKLLKVAFAAIVVSGSALAGVVPQWGGGNGHNGGSSGPDSTLSIYQYGSANAALQ 60

QY 61 LVTRVVTHEMAHAGYNGGADVGQADNSTIELTONGFRNNATIDOWNAKNSDITVGYGG 120  
 DB 61 SDARKSETTITQSGYNGGADVGQADNSTIELTONGFRNNATIDOWNAKNSDITVGYGG 120

QY 121 NNAALVNOTASDSSVMVRQVFGNNATANQY 151  
 DB 121 NNAALVNOTASDSSVMVRQVFGNNATANQY 151

RESULT 12  
 AAB36355  
 ID AAB36355 standard; protein; 151 AA.  
 AC AAB36355;  
 XX  
 DT 26-FEB-2001 (first entry)  
 XX  
 DE Agfa::PT3#10 amino acid sequence SEQ ID NO:30.  
 XX  
 KW Salmonella; agfa; chromosomal gene replacement; fimbrin; epitope;  
 XX vaccine; immune response; immunogen.  
 XX  
 OS Salmonella enteritidis.  
 OS Escherichia coli.  
 OS Synthetic.  
 XX  
 PN WO200060102-A2.  
 XX  
 PD 12-OCT-2000.  
 XX  
 PF 05-APR-2000; 2000WO-CA000356.  
 XX  
 PR 05-APR-1999; 99US-0127888P.  
 XX  
 PA (UYVI-) UNIV VICTORIA.  
 XX  
 PI White AP, Doran JL, Collison SK, Kay WW;  
 XX  
 DR WPI; 2000-672631/65.  
 XX  
 DR N-PSDB; AAC64631.  
 XX  
 PT Recombinant agfa gene having a segment replaced by a foreign DNA sequence  
 PT which encodes foreign epitope or antigen, expresses recombinant Agfa  
 PT protein useful for eliciting immune response in animal.  
 XX  
 PS Disclosure; Page 139; 139pp; English.

The present invention describes a recombinant agfa gene (I) where a segment of the gene has been replaced by a segment of a foreign DNA sequence which encodes a foreign epitope or antigen. Also described are: (1) use of thin aggregative fimbriae (SEF17/TAF) nucleation depended assembly system of strains of Salmonella, Escherichia coli and Enterobacteriaceae for the production of fimbriae comprising recombinant Agfa, CsgA and Agfa-homologue fimbrial subunits, respectively; (2) directing recombination of a recombinant gene into the chromosome of the homologous species; (3) directing recombination of a recombinant gene back into the chromosome of the homologous species, replacing the native copy of that gene; and (4) eliciting an immune response in an animal, comprising separating an amino acid polymer comprising a recombinant Agfa protein containing a replacement segment or segments of foreign amino acid sequence or sequences grown on a Salmonella, E. coli or

CC (1) use of thin aggregative fimbriae (SEF17/TAF) nucleation depended  
 CC assembly system of strains of Salmonella, Escherichia coli and  
 CC Enterobacteriaceae for the production of fimbriae comprising recombinant  
 CC AgfA, CsgA and AgfA-homologue fimbrial subunits, respectively; (2)  
 CC directing recombination of a recombinant gene into the chromosome of the  
 CC homologous species; (3) directing recombination of a recombinant gene  
 CC back into the chromosome of the homologous species, replacing the native  
 CC copy of that gene; and (4) eliciting an immune response in an animal,  
 CC comprising separating an amino acid polymer comprising a recombinant AgfA  
 CC protein containing a replacement segment or segments of foreign amino  
 CC acid sequence or sequences grown on a Salmonella, E. coli or  
 CC Enterobacteriaceae host cell, from the host cell and introducing the  
 CC polymer into the animal in conjunction with a carrier or diluent. (I) is  
 CC useful for the expression of recombinant AgfA protein which is useful for  
 CC eliciting an immune response in an animal. In a fimbrial presentation  
 CC system the heterologous antigens are presented in high numbers (up to  
 CC 500,000 copies/cell), the hybrid fimbrial protein possesses both the  
 CC immunogenicity and adhesion properties relevant for an efficient live  
 CC vaccine, the carrier fimbrial subunit proteins are usually strong  
 CC immunogens, which may be important for directing an immune response  
 CC against the inserted epitope, and hybrid fimbriae are easy and  
 CC inexpensive to purify in large amount. The present sequence is given in  
 CC the exemplification of the present invention

XX SQ Sequence 151 AA;

Query Match 76.7%; Score 600; DB 3; Length 151;  
 Best Local Similarity 81.5%; Pred. No. 7.8e-50;  
 Matches 123; Conservative 4; Mismatches 24; Indels 0; Gaps 0;  
 QY 1 MKLLKVAAPAAIIVSGSALAGVVPQWGGGNGGNSGGPSTLSIYQYGSANAALYDQ 60  
 DB 1 MKLLKVAAPAAIIVSGSALAGVVPQWGGGNGGNSGGPSTLSIYQYGSANAALYDQ 60  
 QY 61 LVTRVVTHEMAHAGYNGADVGQGDNSTIETQNGFRNNATIDOWNAKNSDITVQYGG 120  
 DB 61 SDARKSETTITQSGYNGADVGQGDNSTIETQNGFRNNATIDOWNAKNSDITVQYGG 120  
 QY 121 NNAALVNQTASDSSVMVRQVGFNNATANQY 151  
 DB 121 NNAALVNQTASDSSVMVRQVGFNNATANQY 151

RESULT 13

AAB36348  
 ID AAB36348 standard; protein; 151 AA.

AC AAB36348;

DT 26-FEB-2001 (first entry)

DE AgfA::PT3#3 amino acid sequence SEQ ID NO:16.

KW Salmonella; agfA; chromosomal gene replacement; fimbrial; epitope;  
 KW vaccine; immune response; immunogen.

OS Salmonella enteritidis.  
 OS Escherichia coli.  
 OS Synthetic.

PN WO200060102-A2.

PD 12-OCT-2000.

PF 05-APR-2000; 2000WO-CA000356.

PR 05-APR-1999; 99US-0127888P.

PA (UYVI-) UNIV VICTORIA.

XX White AP, Doran JL, Collison SK, Kay WW;

DR WPI; 2000-672631/65.

DR N-PSDB; AAC64624.

XX Recombinant agfA gene having a segment replaced by a foreign DNA sequence  
 PT which encodes foreign epitope or antigen, expresses recombinant AgfA  
 PT protein useful for eliciting immune response in animal.

PS Disclosure; Page 136; 139pp; English.

XX The present invention describes a recombinant agfA gene (I) where a  
 CC segment of the gene has been replaced by a segment of a foreign DNA  
 CC sequence which encodes a foreign epitope or antigen. Also described are:  
 CC (1) use of thin aggregative fimbriae (SEF17/TAF) nucleation depended  
 CC assembly system of strains of Salmonella, Escherichia coli and  
 CC Enterobacteriaceae for the production of fimbriae comprising recombinant  
 CC AgfA, CsgA and AgfA-homologue fimbrial subunits, respectively; (2)  
 CC directing recombination of a recombinant gene into the chromosome of the  
 CC homologous species; (3) directing recombination of a recombinant gene  
 CC back into the chromosome of the homologous species, replacing the native  
 CC copy of that gene; and (4) eliciting an immune response in an animal,  
 CC comprising separating an amino acid polymer comprising a recombinant AgfA  
 CC protein containing a replacement segment or segments of foreign amino  
 CC acid sequence or sequences grown on a Salmonella, E. coli or  
 CC Enterobacteriaceae host cell, from the host cell and introducing the  
 CC polymer into the animal in conjunction with a carrier or diluent. (I) is  
 CC useful for the expression of recombinant AgfA protein which is useful for  
 CC eliciting an immune response in an animal. In a fimbrial presentation  
 CC system the heterologous antigens are presented in high numbers (up to  
 CC 500,000 copies/cell), the hybrid fimbrial protein possesses both the  
 CC immunogenicity and adhesion properties relevant for an efficient live  
 CC vaccine, the carrier fimbrial subunit proteins are usually strong  
 CC immunogens, which may be important for directing an immune response  
 CC against the inserted epitope, and hybrid fimbriae are easy and  
 CC inexpensive to purify in large amount. The present sequence is given in  
 CC the exemplification of the present invention

XX SQ Sequence 151 AA;

Query Match 73.8%; Score 577; DB 3; Length 151;  
 Best Local Similarity 80.8%; Pred. No. 1.3e-47;  
 Matches 122; Conservative 5; Mismatches 24; Indels 0; Gaps 0;

QY 1 MKLLKVAAPAAIIVSGSALAGVVPQWGGGNGGNSGGPSTLSIYQYGSANAALYDQ 60  
 DB 1 MKLLKVAAPAAIIVSGSALAGVVPQWGGGNGGNSGGPSTLSIYQYGSANAALYDQ 60

QY 61 LVTRVVTHEMAHAGYNGADVGQGDNSTIETQNGFRNNATIDOWNAKNSDITVQYGG 120  
 DB 61 SDARKSETTITQSGYNGADVGQGDNSTIETQNGFRNNATIDOWNAKNSDITVQYGG 120

QY 121 NNAALVNQTASDSSVMVRQVGFNNATANQY 151

DB 121 NNAALVNQTASDSSVMVRQVGFNNATANQY 151

RESULT 14

AAB36343  
 ID AAB36343 standard; protein; 151 AA.

AC AAB36343;

DT 26-FEB-2001 (first entry)

DE Escherichia coli CsgA amino acid sequence SEQ ID NO:7.

KW Salmonella; agfA; chromosomal gene replacement; fimbrial; epitope;  
 KW vaccine; immune response; immunogen.

OS Escherichia coli.

PN WO200060102-A2.

XX 12-OCT-2000.

PF 05-APR-2000; 2000WO-CA000356.  
XX 05-APR-1999; 99US-0127888P.  
XX (UYVI-) UNIV VICTORIA.  
XX White AP, Doran JL, Collison SK, Kay WW;  
XX WPI; 2000-672631/65.  
XX N-PSDB; AAC64619.  
XX Recombinant agfa gene having a segment replaced by a foreign DNA sequence  
XX PT which encodes foreign epitope or antigen, expresses recombinant Agfa  
XX PT protein useful for eliciting immune response in animal.  
XX  
XX Disclosure; Page 135; 139pp; English.  
XX  
XX The present invention describes a recombinant agfa gene (I) where a  
XX segment of the gene has been replaced by a segment of a foreign DNA  
XX sequence which encodes a foreign epitope or antigen. Also described are:  
XX (1) use of thin aggregative fimbriae (SEF17/TAF) nucleation depended  
XX assembly system of strains of Salmonella, Escherichia coli and  
XX Enterobacteriaceae for the production of fimbriae comprising recombinant  
XX Agfa, CsgA and Agfa-homologue fimbrial subunits, respectively; (2)  
XX directing recombination of a recombinant gene into the chromosome of the  
XX homologous species; (3) directing recombination of a recombinant gene  
XX back into the chromosome of the homologous species, replacing the native  
XX copy of that gene; and (4) eliciting an immune response in an animal,  
XX comprising separating an amino acid polymer comprising a recombinant Agfa  
XX protein containing a replacement segment or segments of foreign amino  
XX acid sequence or sequences grown on a Salmonella, E. coli or  
XX Enterobacteriaceae host cell, from the host cell and introducing the  
XX polymer into the animal in conjunction with a carrier or diluent. (I) is  
XX useful for the expression of recombinant Agfa protein which is useful for  
XX eliciting an immune response in an animal. In a fimbrial presentation  
XX system the heterologous antigens are presented in high numbers (up to  
XX 500,000 copies/cell), the hybrid fimbrial protein possesses both the  
XX immunogenicity and adhesion properties relevant for an efficient live  
XX vaccine, the carrier fimbrial subunit proteins are usually strong  
XX immunogens, which may be important for directing an immune response  
XX against the inserted epitope, and hybrid fimbriae are easy and  
XX inexpensive to purify in large amount. The present sequence is given in  
XX the exemplification of the present invention  
XX  
XX Sequence 151 AA;  
XX  
XX Query Match 66.9%; Score 523; DB 3; Length 151;  
XX Best Local Similarity 68.9%; Pred. No. 2e-42;  
XX Matches 104; Conservative 18; Mismatches 29; Indels 0; Gaps 0;  
XX  
XX 1 MKLLKVAFAAIVVSGSALAGVVPQGGGNGHGGNSGPDSTLSIYQYGSANAALYDQ 60  
XX 1 MKLLKVAATAAIVFSGSALAGVVPQYGGGNGHGGNSGPNSELNIYQYGGNSALALQ 60  
XX  
XX 61 LVTRVVTHEMAHAGYNGADVGGADNSTLTQTNGFRNNTATIDQWNAKNSDITVGOYGG 120  
XX 61 TDARNSDLITQHGCGNGADVGGQSDSSIDLTLQRFNGSATLDQWNGKNSMTVKQFGG 120  
XX  
XX 121 NNAALVNQTSADSSVMYRVQVGFNGNATANQY 151  
XX 121 GNGRAVDQTASNSVNVTVQVGFNGNATAHQY 151  
XX  
XX Search completed: August 2, 2004, 14:48:28  
XX Job time : 45.9 secs

KW Plasma protein; immune response; antibacterial; vaccine; gene therapy.  
XX Escherichia coli.  
XX WO2003064446-A2.  
XX 07-AUG-2003.  
XX 30-JAN-2003; 2003WO-EP000943.  
XX 31-JAN-2002; 2002GB-00002275.  
XX (HANS-) HANSA MEDICAL RES AB.  
XX Bjoerck L, Olsen A, Wikstroem M, Herwald H;  
XX WPI; 2003-646136/61.  
XX N-PSDB; ACP36153.  
XX  
XX New isolated peptide capable of binding a mammalian plasma protein,  
XX useful in the manufacture of a medicament for the prevention and/or  
XX treatment of a bacterial infection, such as Escherichia coli, Salmonella  
XX or Shigella infections.  
XX  
XX Disclosure; Page 41-42; 42pp; English.  
XX  
XX The invention relates to an isolated peptide capable of binding a  
XX mammalian plasma protein or of generating an immune response in a mammal  
XX selected from sequences shown in ABR82642, ABR82648-49. The peptide or  
XX antibody is useful for treating a bacterial infection in a human or  
XX animal or in the manufacture of a medicament for the prophylactic  
XX treatment of a bacterial infection, such as Escherichia coli, Salmonella  
XX or Shigella infection. The peptide that is immobilized on a solid support  
XX is also useful as a reagent for determining the ability of a plasma  
XX protein to bind to bacteria. The present sequence represents an E. coli  
XX 15 kDa protein  
XX  
XX Sequence 151 AA;  
XX  
XX Query Match 66.2%; Score 518; DB 7; Length 151;  
XX Best Local Similarity 68.2%; Pred. No. 6.2e-42;  
XX Matches 103; Conservative 18; Mismatches 30; Indels 0; Gaps 0;  
XX  
XX 1 MKLLKVAFAAIVVSGSALAGVVPQGGGNGHGGNSGPDSTLSIYQYGSANAALYDQ 60  
XX 1 MKLLKVAATAAIVFSGSALAGVVPQYGGGNGHGGNSGPNSELNIYQYGGNSALALQ 60  
XX  
XX 61 LVTRVVTHEMAHAGYNGADVGGADNSTLTQTNGFRNNTATIDQWNAKNSDITVGOYGG 120  
XX 61 TDARNSDLITQHGCGNGADVGGQSDSSIDLTLQRFNGSATLDQWNGKNSMTVKQFGG 120  
XX  
XX 121 NNAALVNQTSADSSVMYRVQVGFNGNATANQY 151  
XX 121 GNGRAVDQTASNSVNVTVQVGFNGNATAHQY 151  
XX  
XX Search completed: August 2, 2004, 14:48:28  
XX Job time : 45.9 secs

This Page Blank (uspto)



GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: August 2, 2004, 14:40:48 ; Search time 12 Seconds  
(without alignments)  
649.627 Million cell updates/sec

Title: US-09-543-407-26  
Perfect score: 782  
Sequence: 1 MKLLKVAFAAIVVSGSALA.....DSSVMVQVGFNNATANYQY 151

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 389414 seqs, 51625971 residues

Total number of hits satisfying chosen parameters: 389414

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Issued Patents AA:\*  
1: /cgn2\_6/ptodata/2/iaa/5A COMB.pep:\*  
2: /cgn2\_6/ptodata/2/iaa/5B COMB.pep:\*  
3: /cgn2\_6/ptodata/2/iaa/6A COMB.pep:\*  
4: /cgn2\_6/ptodata/2/iaa/6B COMB.pep:\*  
5: /cgn2\_6/ptodata/2/iaa/PCTUS COMB.pep:\*  
6: /cgn2\_6/ptodata/2/iaa/backfiles1.pep:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	687	87.9	151	1	US-08-233-788A-59
2	507	64.8	120	1	US-08-233-788A-57
3	100	12.8	943	4	US-09-056-556-204
4	100	12.8	943	4	US-09-072-596-199
5	100	12.8	943	4	US-09-477-135A-131
6	100	12.8	943	4	US-09-072-967-204
7	98.5	12.6	738	3	US-08-864-038A-3
8	93	11.9	892	4	US-09-336-447A-5
9	91.5	11.7	975	4	US-09-328-352-4764
10	88	11.3	273	4	US-09-328-352-6167
11	86	11.0	1415	4	US-09-252-991A-26438
12	82	10.5	673	3	US-09-196-387-8
13	82	10.5	673	3	US-09-841-835-8
14	82	10.5	949	3	US-09-196-387-10
15	82	10.5	949	3	US-09-841-835-10
16	82	10.5	1327	3	US-09-196-387-2
17	82	10.5	1327	4	US-09-841-835-2
18	82	10.5	1327	4	US-09-372-115A-8
19	81.5	10.4	873	4	US-09-336-447A-13
20	81	10.4	2736	4	US-09-252-991A-30227
21	80	10.2	863	4	US-08-489-039A-10179
22	79	10.1	353	2	US-08-887-702-37
23	78.5	10.0	745	4	US-09-336-115C-6
24	78.5	10.0	1739	4	US-09-540-236-3739
25	78	10.0	1216	4	US-09-434-000C-5130
26	77.5	9.9	321	4	US-09-498-520A-18
27	77.5	9.9	713	4	US-09-059-584-53

28	77	9.8	1338	2	US-08-728-470-9	Sequence 9, Appli
29	77	9.8	1338	3	US-08-719-641-9	Sequence 9, Appli
30	77	9.8	1338	2	US-08-728-470-10	Sequence 10, Appl
31	77	9.8	1529	3	US-08-719-641-10	Sequence 10, Appl
32	77	9.8	1599	2	US-08-617-697-9	Sequence 9, Appli
33	77	9.8	1600	2	US-08-617-697-10	Sequence 10, Appl
34	77	9.8	2315	4	US-09-543-681A-5434	Sequence 5434, Ap
35	76.5	9.8	363	1	US-08-458-023B-6	Sequence 6, Appli
36	76.5	9.8	435	2	US-08-331-515A-2	Sequence 2, Appli
37	76.5	9.8	435	3	US-09-168-408A-2	Sequence 2, Appli
38	76.5	9.8	941	4	US-09-336-447A-9	Sequence 2, Appli
39	76.5	9.8	1612	1	US-08-169-927-2	Sequence 2, Appli
40	76	9.7	364	1	US-07-792-259-17	Sequence 17, Appl
41	76	9.7	1385	1	US-07-876-280-2	Sequence 2, Appli
42	76	9.7	1385	1	US-07-675-772-2	Sequence 2, Appli
43	76	9.7	1385	1	US-08-063-170-2	Sequence 2, Appli
44	76	9.7	1385	1	US-08-158-232-2	Sequence 2, Appli
45	76	9.7	1385	1	US-08-304-626-2	Sequence 2, Appli

ALIGNMENTS

RESULT 1  
US-08-233-788A-59  
; Sequence 59, Application US/08233788A  
; Patent No. 5635617  
; GENERAL INFORMATION:  
; APPLICANT: Doran, James L.  
; APPLICANT: Kay, William W.  
; APPLICANT: Collinson, Karen S.  
; APPLICANT: Clouthier, Sharon C.  
; TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR DETECTION  
; TITLE OF INVENTION: OF SALMONELLA  
; NUMBER OF SEQUENCES: 61  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Seed and Berry  
; STREET: 6300 Columbia Center, 701 Fifth Avenue  
; CITY: Seattle  
; STATE: Washington  
; COUNTRY: U.S.A.  
; ZIP: 98104-7092  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/233,788A  
; FILING DATE: 26-APR-1994  
; CLASSIFICATION: 435  
; ATTORNEY/AGENT INFORMATION:  
; NAME: King, Joshua  
; REGISTRATION NUMBER: 35,570  
; REFERENCE/DOCKET NUMBER: 920043.403C2  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (206) 622-4900  
; TELEFAX: (206) 682-6031  
; TELEX: 3723836 SEEDANBERRY  
; INFORMATION FOR SEQ ID NO: 59:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 151 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
; US-08-233-788A-59

Query Match 87.9%; Score 687; DB 1; Length 151;  
Best Local Similarity 90.1%; Pred. No. 7e-60;  
Matches 136; Conservative 13; Mismatches 13; Indels 0; Gaps 0;

1 MKLLKVAFAAIVVSGSALAAGVQVGGGNGHNGSGSPDSTLSIYQVGSANAALYDQ 60

Db 1 MKLLKVAFAAIVVSGSALAGVVPQWGGGNGHGGSSGPDSTLSIYQGSANAALALQ 60  
 QY 61 LVTRVVTTHSMAGYNGADVGCGADNSTIETQNGFRNATIDOWNAKNSDITVGOYGG 120  
 Db 61 SARKESTTITSGYNGADVGCGADNSTIETQNGFRNATIDOWNAKNSDITVGOYGG 120  
 QY 121 NNAALVNQTSDDSSVMVRQVGFNNATANQY 151  
 Db 121 NNPALVNQTSDDSSVMVRQVGFNNATANQY 151

## RESULT 2

US-08-233-788A-57  
 ; Sequence 57, Application US/08233788A  
 ; Patent No. 5635617

## GENERAL INFORMATION:

APPLICANT: Doran, James L.  
 APPLICANT: Kay, William W.  
 APPLICANT: Collinson, Karen S.  
 APPLICANT: Clouthier, Sharon C.  
 TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR DETECTION  
 OF SALMONELLA  
 NUMBER OF SEQUENCES: 61  
 CORRESPONDENCE ADDRESS:  
 ADDRESSEE: Seed and Berry  
 STREET: 6300 Columbia Center, 701 Fifth Avenue  
 CITY: Seattle  
 STATE: Washington  
 COUNTRY: U.S.A.  
 ZIP: 98104-7092

## COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk  
 COMPUTER: IBM PC compatible  
 OPERATING SYSTEM: PC-DOS/MS-DOS  
 SOFTWARE: Patent In Release #1.0, Version #1.25  
 CURRENT APPLICATION DATA:  
 APPLICATION NUMBER: US/08/233,788A  
 FILING DATE: 26-APR-1994  
 CLASSIFICATION: 435

## ATTORNEY/AGENT INFORMATION:

NAME: King, Joshua  
 REGISTRATION NUMBER: 35,570  
 REFERENCE/DOCKET NUMBER: 920043.403C2  
 TELECOMMUNICATION INFORMATION:  
 TELEPHONE: (206) 622-4900  
 TELEFAX: (206) 682-6031  
 TELEX: 3723836 SEEDANBERRY  
 INFORMATION FOR SEQ ID NO: 57:  
 SEQUENCE CHARACTERISTICS:  
 LENGTH: 120 amino acids  
 TYPE: amino acid  
 TOPOLOGY: linear

## MOLECULE TYPE:

protein

## US-08-233-788A-57

Query Match 64.8%; Score 507; DB 1; Length 120;  
 Best Local Similarity 87.5%; Pred. No. 2.1e-42;  
 Matches 98; Conservative 2; Mismatches 12; Indels 0; Gaps 0;  
 QY 22 VYPQGGGNGHGGSSGPDSTLSIYQGSANAALALQSDARKSETTITQSGYNGADV 81  
 Db 1 VYPQGGGNGHGGSSGPDSTLSIYQGSANAALALQSDARKSETTITQSGYNGADV 60  
 QY 82 GCGADNSTIETQNGFRNATIDOWNAKNSDITVGOYGNNAALVNQTSADS 133  
 Db 61 GCGADNSTIETQNGFRNATIDOWNAKNSDITVGOYGNNAALVNQTSADS 112

## RESULT 3

US-09-556-556-204  
 ; Sequence 204, Application US/09056556  
 ; Patent No. 6350456

## GENERAL INFORMATION:

APPLICANT: Reed, Steven G.  
 APPLICANT: Skeiky, Yasir A.W.  
 APPLICANT: Dillon, Davin C.  
 TITLE OF INVENTION: COMPOUNDS AND METHODS FOR THE PREVENTION AND  
 TREATME  
 NUMBER OF SEQUENCES: 241  
 CORRESPONDENCE ADDRESS:  
 ADDRESSEE: SEED AND BERRY LLP  
 STREET: 6300 Columbia Center, 701 Fifth Avenue  
 CITY: Seattle  
 STATE: Washington  
 COUNTRY: USA  
 ZIP: 98104-7092

## COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk  
 COMPUTER: IBM PC compatible  
 OPERATING SYSTEM: PC-DOS/MS-DOS  
 SOFTWARE: Patent In Release #1.0, Version #1.30  
 CURRENT APPLICATION DATA:  
 APPLICATION NUMBER: US/09/056,556  
 FILING DATE: 07-APR-1998  
 CLASSIFICATION:

## ATTORNEY/AGENT INFORMATION:

NAME: Maki, David J.  
 REGISTRATION NUMBER: 31,392  
 REFERENCE/DOCKET NUMBER: 210121.457  
 TELECOMMUNICATION INFORMATION:  
 TELEPHONE: (206) 622-4900  
 TELEFAX: (206) 682-6031  
 INFORMATION FOR SEQ ID NO: 204:  
 SEQUENCE CHARACTERISTICS:  
 LENGTH: 943 amino acids  
 TYPE: amino acid  
 STRANDEDNESS:  
 TOPOLOGY: linear

## US-09-056-556-204

Query Match 12.8%; Score 100; DB 4; Length 943;  
 Best Local Similarity 26.0%; Pred. No. 0.2; Indels 50; Gaps 7;  
 Matches 40; Conservative 12; Mismatches 52  
 QY 16 GSALAGVVPQWGGG-GNHN-GGNSGSPDSTLSIYQGSANAALYDQLVTRVVTTHMAHA 73  
 Db 464 GSGNIGVFNVGSSGLGNYNIGSN-----LGIYNIGFNGVDY-----NV 503  
 QY 74 GYNGADVGCGADNSTIETQNGFRNATIDOWNAKNSDITVGOYGGN-----121  
 Db 504 GFGNAGDFNQGFANT-----GNNNIGFANTNNNIGLSDGDNQOQGFNIASGWS 553  
 QY 122 ---NAALVNQTSDDSSVM---VRQVGFNNATAN 149  
 Db 554 GTGNSGLFNSGNNVGVIPNAGTGNVGIANSIGTN 587

## RESULT 4

US-09-072-596-199  
 ; Sequence 199, Application US/09072596  
 ; Patent No. 6458366

## GENERAL INFORMATION:

APPLICANT: Reed, Steven G.  
 APPLICANT: Skeiky, Yasir A.W.  
 APPLICANT: Dillon, Davin C.  
 APPLICANT: Campos-Neto, Antonia  
 APPLICANT: Houghton, Raymond  
 APPLICANT: Vedwick, Thomas S.  
 APPLICANT: Twardzik, Daniel R.  
 APPLICANT: Lodes, Michael J.  
 APPLICANT: Hendrickson, Ronald C.

## TITLE OF INVENTION:

COMPOUNDS AND METHODS FOR DIAGNOSIS OF

TUBERCULOSIS

NUMBER OF SEQUENCES: 350

CORRESPONDENCE ADDRESS:

ADDRESSEE: SEED AND BERRY LLP

STREET: 6300 Columbia Center, 701 Fifth Avenue

CITY: Seattle

```
STATE: Washington
COUNTRY: USA
ZIP: 98104-7092
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/072,596
FILING DATE: 05-MAY-1998
CLASSIFICATION:
ATTORNEY/AGENT INFORMATION:
NAME: Maki, David J.
REGISTRATION NUMBER: 31,392
REFERENCE/DOCKET NUMBER: 210121.417C9
TELEPHONE: (206) 622-4900
TELEFAX: (206) 682-6031
INFORMATION FOR SEQ ID NO: 199:
SEQUENCE CHARACTERISTICS:
LENGTH: 943 amino acids
TYPE: amino acid
STRANDEDNESS:
TOPOLOGY: linear
US-09-072-596-199

Query Match 12.8%; Score 100; DB 4; Length 943;
Best Local Similarity 26.0%; Pred. No. 0.2;
Matches 40; Conservative 12; Mismatches 52; Indels 50; Gaps 7;

QY 16 GSALAGVVPQGGG-GNHN-GGNSGGPDSTLSIYQYGSANAALYDQLVTRVVTHEMAHA 73
Db 571 GSGNIGVFNVGSGSLGNVIGSGN-----LGIYNIGFNGVDY-----NV 610

QY 74 GYGNAGADVCGADNSTIETQNGFRNATIDOWNAKNSDITVQYGGN-----121
Db 611 GFGNAGDFNQGFANT-----GNNNIGFANTGNNNIGLIGSGDNQOQGFNIASGWN 660

QY 122 ---NAALVNQATSDSSVM---VRQVGFGNATAN 149
Db 661 GTGNSGLFNSGNTNNVGIFNAGTGNVGIANSGTGN 694

RESULT 6
US-09-072-967-204
Sequence 204, Application US/09072967
Patent No. 6592877
GENERAL INFORMATION:
APPLICANT: Reed, Steven G.
APPLICANT: Skeiky, Yasir A.W.
APPLICANT: Dillon, David C.
APPLICANT: Campos-Neto, Antonio
APPLICANT: Houghton, Raymond
APPLICANT: Vedwick, Thomas S.
APPLICANT: Twardzik, Daniel R.
APPLICANT: Lodes, Michael J.
APPLICANT: Hendrickson, Ronald C.
TITLE OF INVENTION: COMPOUNDS AND METHODS FOR IMMUNOTHERAPY
TITLE OF INVENTION: AND DIAGNOSIS OF TUBERCULOSIS
NUMBER OF SEQUENCES: 355
CORRESPONDENCE ADDRESS:
ADDRESSEE: SEED and BERRY LLP
STREET: 6300 Columbia Center, 701 Fifth Avenue
CITY: Seattle
STATE: Washington
COUNTRY: USA
ZIP: 98104-7092
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/072,967
FILING DATE: 05-MAY-1998
CLASSIFICATION:
ATTORNEY/AGENT INFORMATION:
NAME: Maki, David J.
REGISTRATION NUMBER: 31,392
REFERENCE/DOCKET NUMBER: 210121.411C9
TELEPHONE: (206) 622-4900
TELEFAX: (206) 682-6031
INFORMATION FOR SEQ ID NO: 204:
SEQUENCE CHARACTERISTICS:
LENGTH: 943 amino acids
TYPE: amino acid
STRANDEDNESS:
TOPOLOGY: linear
US-09-072-967-204

Query Match 12.8%; Score 100; DB 4; Length 943;
Best Local Similarity 26.0%; Pred. No. 0.2;
Matches 40; Conservative 12; Mismatches 52; Indels 50; Gaps 7;

QY 16 GSALAGVVPQGGG-GNHN-GGNSGGPDSTLSIYQYGSANAALYDQLVTRVVTHEMAHA 73
Db 464 GSGNIGVFNVGSGSLGNVIGSGN-----LGIYNIGFNGVDY-----NV 503

QY 74 GYGNAGADVCGADNSTIETQNGFRNATIDOWNAKNSDITVQYGGN-----121
Db 504 GFGNAGDFNQGFANT-----GNNNIGFANTGNNNIGLIGSGDNQOQGFNIASGWN 553

STATE: Washington
COUNTRY: USA
ZIP: 98104-7092
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/477,135A
FILING DATE: 2000-01-03
CURRENT FILING DATE: 1997-12-15
PRIOR APPLICATION NUMBER: 08990823
PRIOR FILING DATE: 1996-06-14
PRIOR APPLICATION NUMBER: 60/000,254
PRIOR FILING DATE: 1995-06-15
NUMBER OF SEQ ID NOS: 169
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 131
LENGTH: 943
TYPE: PRT
ORGANISM: Mycobacterium tuberculosis
US-09-477-135A-131

Query Match 12.8%; Score 100; DB 4; Length 943;
Best Local Similarity 26.0%; Pred. No. 0.2;
Matches 40; Conservative 12; Mismatches 52; Indels 50; Gaps 7;

QY 16 GSALAGVVPQGGG-GNHN-GGNSGGPDSTLSIYQYGSANAALYDQLVTRVVTHEMAHA 73
Db 464 GSGNIGVFNVGSGSLGNVIGSGN-----LGIYNIGFNGVDY-----NV 503

QY 74 GYGNAGADVCGADNSTIETQNGFRNATIDOWNAKNSDITVQYGGN-----121
Db 504 GFGNAGDFNQGFANT-----GNNNIGFANTGNNNIGLIGSGDNQOQGFNIASGWN 553

STATE: Washington
COUNTRY: USA
ZIP: 98104-7092
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/477,135A
FILING DATE: 2000-01-03
CURRENT FILING DATE: 1997-12-15
PRIOR APPLICATION NUMBER: 08990823
PRIOR FILING DATE: 1996-06-14
PRIOR APPLICATION NUMBER: 60/000,254
PRIOR FILING DATE: 1995-06-15
NUMBER OF SEQ ID NOS: 169
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 131
LENGTH: 943
TYPE: PRT
ORGANISM: Mycobacterium tuberculosis
US-09-477-135A-131

Query Match 12.8%; Score 100; DB 4; Length 943;
Best Local Similarity 26.0%; Pred. No. 0.2;
Matches 40; Conservative 12; Mismatches 52; Indels 50; Gaps 7;
```



Db 393 ASNGEHNYGIGN 407

RESULT 10

US-09-328-352-6167

Sequence 6167, Application US/09328352

Patent No. 6562958

GENERAL INFORMATION:

APPLICANT: Gary L. Breton et al.

TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO ACINETOBACTER

FILE OF INVENTION: BAUMANNII FOR DIAGNOSTICS AND THERAPEUTICS

FILE REFERENCE: GTC99-03PA

CURRENT APPLICATION NUMBER: US/09/328,352

CURRENT FILING DATE: 1999-06-04

NUMBER OF SEQ ID NOS: 8252

SEQ ID NO 6167

LENGTH: 273

TYPE: PRT

ORGANISM: Acinetobacter baumannii

US-09-328-352-6167

Query Match 11.3%; Score 88; DB 4; Length 273;

Best Local Similarity 28.5%; Pred. No. 0.6;

Matches 43; Conservative 19; Mismatches 65; Indels 24; Gaps 7;

QY 1 MKLLKVA--FAAIVVSGSALAGVVPQWGGGNH----NGGNGSS-GPDSTLSIYQYGA 53

Db 19 MKKLAISALLSALVSGAANA-YQAEVGGSYNYLDPDNGSSVSKFGYDGYTFYFVQTR 77

QY 54 NAALYDQLVTRVVTHEMAHAGYNGADVGQGDNSTIELTQNGFRNATIDQWAKNSDI 113

Db 78 NAPLAEAFLNPRASNVNHFVNYG-----DNSGTDQYGVG-----VEYFVPSDF 123

QY 114 TVGQYGGNNAALVNOTASDSSV--MVROVGF 142

Db 124 YLSDGVGRNEREIDNINIDSKVTYAAEVGY 154

RESULT 11

US-09-252-991A-26438

Sequence 26438, Application US/09252991A

Patent No. 6551795

GENERAL INFORMATION:

APPLICANT: Marc J. Rubenfield et al.

TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS

FILE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS

FILE REFERENCE: 107196.136

CURRENT APPLICATION NUMBER: US/09/252,991A

CURRENT FILING DATE: 1999-02-18

PRIOR APPLICATION NUMBER: US 60/074,788

PRIOR FILING DATE: 1998-02-18

PRIOR APPLICATION NUMBER: US 60/094,190

PRIOR FILING DATE: 1998-07-27

NUMBER OF SEQ ID NOS: 33142

SEQ ID NO 26438

LENGTH: 1415

TYPE: PRT

ORGANISM: Pseudomonas aeruginosa

US-09-252-991A-26438

Query Match 11.0%; Score 86; DB 4; Length 1415;

Best Local Similarity 23.9%; Pred. No. 7.8;

Matches 38; Conservative 21; Mismatches 68; Indels 32; Gaps 5;

QY 13 VVSGSALAGVVPQWGGGNH-----GNSSGPDSTLSIYQYGAANAALYDQLVT 63

Db 366 ILSHEVSVAAVQANAAGDGGQVHVQAQFAGANASDSNGVTIVQQQPAVDLAAGANGTS 425

QY 64 RVVTHEMAHAGY-----NGADVGGADNSTIELTON-----GFRNATIDQW 107

Db 426 AVQSGGANIGSGANGISVVQSGNAGANIGAGASDISVQSGSNIGSGVGVTVVQSQN 485

QY 108 AKN-----SDITVQYGGNNAALVNOTASDSSVMVRQVG 141

Db 486 GANIGSGASGITVQ--SONGANIGSGASGISVWQSQSG 522

RESULT 12

US-09-196-387-8

Sequence 8, Application US/09196387

Patent No. 6277613

GENERAL INFORMATION:

APPLICANT: de Lange, Titia

APPLICANT: Smith, Susan

TITLE OF INVENTION: A PROTEIN THAT BINDS TO TRF1 AND METHODS

FILE OF INVENTION: OF USE THEREOF

NUMBER OF SEQUENCES: 12

CORRESPONDENCE ADDRESS:

ADDRESSEE: Klauber & Jackson

STREET: 411 Hackensack Avenue, 4th Floor

CITY: Hackensack

STATE: New Jersey

COUNTRY: USA

ZIP: 07601

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: PatentIn Release #1.0, Version #1.30

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/09/196,387

FILING DATE:

CLASSIFICATION:

PRIOR APPLICATION DATA:

APPLICATION NUMBER: 09/095,225

FILING DATE: June 10, 1998

ATTORNEY/AGENT INFORMATION:

NAME: Jackson Esq., David A.

REGISTRATION NUMBER: 26,742

REFERENCE/DOCKET NUMBER: 600-1-230 CIP1

TELECOMMUNICATION INFORMATION:

TELEPHONE: 201-487-5800

TELEFAX: 201-343-1684

TELEX: 133521

INFORMATION FOR SEQ ID NO: 8:

SEQUENCE CHARACTERISTICS:

LENGTH: 673 amino acids

TYPE: amino acid

TOPOLOGY: linear

MOLECULE TYPE: protein

US-09-196-387-8

Query Match 10.5%; Score 82; DB 3; Length 673;

Best Local Similarity 29.6%; Pred. No. 7.4;

Matches 34; Conservative 14; Mismatches 59; Indels 8; Gaps 4;

QY 6 VAAFAAI-VVSGSALAGVVPQWGGGNHNGGNSGPDSTLSIYQYGAANAALYDQL--- 61

Db 99 VAAAPVVPVAVSTSSAAGVAPNPAGSGNNSPSSSPTSS--SSSSPSSPGSSLAESPEAA 157

QY 62 -VTRVVTHEMAHAGYNGADVGQGDNSTIELTONG--FRNATIDQWAKNSDI 113

Db 158 GVSSTAPLPGAGGTGTGVPVAVSGALRELLEACRNGDVSRVKRLVDANVNAKDM 212

RESULT 13

US-09-841-835-8

Sequence 8, Application US/09841835

Patent No. 6506587

GENERAL INFORMATION:

APPLICANT: de Lange, Titia

APPLICANT: Smith, Susan

TITLE OF INVENTION: A PROTEIN THAT BINDS TO TRF1 AND METHODS

FILE OF INVENTION: OF USE THEREOF

NUMBER OF SEQUENCES: 12

CORRESPONDENCE ADDRESS:



```

; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-09-841-835-10

Query Match      10.5%; Score 82; DB 4; Length 949;
Best Local Similarity 29.6%; Pred. No. 12;
Matches 34; Conservative 14; Mismatches 59; Indels 8; Gaps 4;

QY 6 VAAFAAI-VWGSALAGVVPQWGGGNNHNGGNSGGPDSTLSIYQYGSANNAALYDQL--- 61
Db 99 VAAAPVPEAVETSSAAGVAPNPAGSGNNPSSSSPTSS-SSSSPSPGSSSLAESPEAA 157

QY 62 -VTRVVTHEMAHAGYNGADYQGGADNSTIELTNG--FRNNATIDOWNAKNSDI 113
Db 158 GVSSTAPLGPGAAGPGTGVPVAVSGALRELLACRNGDVSRVKRLVDAANTVNAKDM 212
```

Search completed: August 2, 2004, 14:58:36  
Job time : 13 secs

This Page Blank (uspto)



GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: August 2, 2004, 14:54:48 ; Search time 36.8 Seconds  
(without alignments)  
1287.123 Million cell updates/sec

Title: US-09-543-407-26

Perfect score: 782

Sequence: 1 MKLLKVAFAAIVVSGSALA.....DSSVMVRQVFGNNATANQY 151

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1291235 seqs, 313682936 residues

Total number of hits satisfying chosen parameters: 1291235

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published Applications AA:

- 1: /cgn2\_6/ptodata/2/pubpaa/US07\_PUBCOMB.pcp.\*
- 2: /cgn2\_6/ptodata/2/pubpaa/PCT\_NEW\_PUB.pcp.\*
- 3: /cgn2\_6/ptodata/2/pubpaa/US05\_NEW\_PUB.pcp.\*
- 4: /cgn2\_6/ptodata/2/pubpaa/US05\_PUBCOMB.pcp.\*
- 5: /cgn2\_6/ptodata/2/pubpaa/US07\_NEW\_PUB.pcp.\*
- 6: /cgn2\_6/ptodata/2/pubpaa/PCTUS\_PUBCOMB.pcp.\*
- 7: /cgn2\_6/ptodata/2/pubpaa/US08\_NEW\_PUB.pcp.\*
- 8: /cgn2\_6/ptodata/2/pubpaa/US08\_PUBCOMB.pcp.\*
- 9: /cgn2\_6/ptodata/2/pubpaa/US09A\_PUBCOMB.pcp.\*
- 10: /cgn2\_6/ptodata/2/pubpaa/US09B\_PUBCOMB.pcp.\*
- 11: /cgn2\_6/ptodata/2/pubpaa/US09C\_PUBCOMB.pcp.\*
- 12: /cgn2\_6/ptodata/2/pubpaa/US09\_NEW\_PUB.pcp.\*
- 13: /cgn2\_6/ptodata/2/pubpaa/US10A\_PUBCOMB.pcp.\*
- 14: /cgn2\_6/ptodata/2/pubpaa/US10B\_PUBCOMB.pcp.\*
- 15: /cgn2\_6/ptodata/2/pubpaa/US10C\_PUBCOMB.pcp.\*
- 16: /cgn2\_6/ptodata/2/pubpaa/US10\_NEW\_PUB.pcp.\*
- 17: /cgn2\_6/ptodata/2/pubpaa/US60\_NEW\_PUB.pcp.\*
- 18: /cgn2\_6/ptodata/2/pubpaa/US60\_PUBCOMB.pcp.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	520	66.5	151	12	US-09-741-873B-4
2	520	66.5	151	12	US-09-741-873B-4
3	442	56.5	131	12	US-09-741-873B-2
4	442	56.5	131	12	US-09-741-873B-2
5	116.5	14.9	445	15	US-10-369-493-20638
6	105	13.4	597	9	US-09-793-306-146
7	100	12.8	943	10	US-09-996-634-131
8	100	12.8	943	10	US-09-997-182-131
9	100	12.8	943	14	US-10-193-002-199
10	100	12.8	943	14	US-10-084-843-204
11	100	12.8	3300	12	US-10-282-122A-64369
12	98.5	12.6	486	15	US-10-369-493-20619
13	98.5	12.6	486	15	US-10-282-122A-49412
14	96	12.3	186	12	US-10-282-122A-49412
15	93	11.9	892	10	US-09-952-267-5

16	92.5	11.8	253	16	US-10-437-963-114193
17	92.5	11.8	1721	12	US-10-282-122A-62548
18	92	11.8	354	12	US-09-820-843A-21
19	92	11.8	616	16	US-10-437-963-193067
20	91.5	11.7	154	16	US-10-437-963-162284
21	91.5	11.7	591	14	US-10-233-553-23
22	91.5	11.7	894	14	US-10-233-553-11
23	91	11.6	678	12	US-10-282-122A-64573
24	91	11.6	974	12	US-10-282-122A-44959
25	91	11.6	1106	12	US-10-282-122A-62472
26	91	11.6	6310	12	US-10-282-122A-67793
27	88.5	11.3	197	12	US-10-425-114-67750
28	88.5	11.3	1621	14	US-10-185-990-10
29	88.5	11.3	1626	14	US-10-185-990-11
30	88	11.3	1448	16	US-10-408-765A-998
31	87.5	11.2	204	12	US-10-424-599-203972
32	87.5	11.2	250	16	US-10-478-670-164
33	87.5	11.2	1649	15	US-10-369-493-18460
34	87	11.1	904	15	US-10-369-493-12420
35	86.5	11.1	2204	12	US-10-282-122A-64364
36	86	11.0	275	12	US-10-424-599-215142
37	85.5	10.9	263	12	US-10-425-114-49960
38	85.5	10.9	278	9	US-09-810-264-28
39	85	10.9	193	16	US-10-437-963-148500
40	84.5	10.8	126	16	US-10-479-670-80
41	84	10.7	191	16	US-10-437-963-105413
42	84	10.7	353	16	US-10-437-963-152921
43	84	10.7	353	16	US-10-437-963-120176
44	84	10.7	391	12	US-10-424-599-148448
45	84	10.7	394	12	US-10-425-114-55337

ALIGNMENTS

RESULT 1

US-09-741-873B-4  
; Sequence 4, Application US/09741873B  
; Publication No. US20020081722A1  
; GENERAL INFORMATION:  
; APPLICANT: Olsen, Arne  
; TITLE OF INVENTION: Fibronectin Binding Protein As Well As Its Preparation  
; FILE REFERENCE: 012889-084  
; CURRENT APPLICATION NUMBER: US/09/741,873B  
; CURRENT FILING DATE: 2003-04-04  
; PRIOR APPLICATION NUMBER: SE 8801723-1  
; PRIOR FILING DATE: 1998-05-06  
; PRIOR APPLICATION NUMBER: US 08/978,878  
; PRIOR FILING DATE: 1997-11-26  
; PRIOR APPLICATION NUMBER: US 07/347,189  
; PRIOR FILING DATE: 1989-05-04  
; PRIOR APPLICATION NUMBER: US 07/789,437  
; PRIOR FILING DATE: 1991-11-06  
; PRIOR APPLICATION NUMBER: US 07/970,846  
; PRIOR FILING DATE: 1992-11-03  
; PRIOR APPLICATION NUMBER: US 08/187,865  
; PRIOR FILING DATE: 1994-01-28  
; PRIOR APPLICATION NUMBER: US 08/318,519  
; PRIOR FILING DATE: 1994-10-05  
; NUMBER OF SEQ ID NOS: 10  
; SOFTWARE: PatentIn version 3.0  
; SEQ ID NO 4  
; LENGTH: 151  
; TYPE: PRT  
; ORGANISM: Escherichia coli  
US-09-741-873B-4

Query Match 66.5%; Score 520; DB 12; Length 151;  
Best Local Similarity 68.2%; Pred. No. 2.2e-44;  
Matches 103; Conservative 19; Mismatches 29; Indels 0; Gaps 0;  
QY 1 MKLLKVAFAAIVVSGSALAGVFPQWGGGNNHGGSSGDPSTLSIYQGSANAALYDQ 60

Db 1 MKLLKVAIAAIVFGSSAVGVVPOYGGGNGHGGGNSGPNSELNIYQYGGNSALALQ 60  
 QY 61 LVTRVVTTHMAHAGYNGADVGGADNNTIELTQNGFNATIDOWNAKNSDITVQYGG 120  
 Db 61 TDARNSDLTITQHGCGGNGADVGGGDDSSIDLTORFGNSATLTDQWNGKNSMTVKQFGG 120  
 QY 121 NNAALVNQTSADSSVVMVQVGFNNATANQY 151  
 Db 121 GNGAAVDQTASNSSVNVTVQVGFNNATAHQY 151

RESULT 2  
 US-09-741-873B-4  
 ; Sequence 4, Application US/09741873B  
 ; Publication No. US20040096965A9  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Normark, Staffan  
 ; APPLICANT: Olsen, Arne  
 ; TITLE OF INVENTION: Fibronectin Binding Protein As Well As Its Preparation  
 ; FILE REFERENCE: 012889-084  
 ; CURRENT APPLICATION NUMBER: US/09/741,873B  
 ; PRIOR FILING DATE: 2003-04-04  
 ; PRIOR APPLICATION NUMBER: SE 8801723-1  
 ; PRIOR FILING DATE: 1998-05-06  
 ; PRIOR APPLICATION NUMBER: US 08/978,878  
 ; PRIOR FILING DATE: 1997-11-26  
 ; PRIOR APPLICATION NUMBER: US 07/347,189  
 ; PRIOR FILING DATE: 1989-05-04  
 ; PRIOR APPLICATION NUMBER: US 07/789,437  
 ; PRIOR FILING DATE: 1991-11-06  
 ; PRIOR APPLICATION NUMBER: US 07/970,846  
 ; PRIOR FILING DATE: 1992-11-03  
 ; PRIOR APPLICATION NUMBER: US 08/187,865  
 ; PRIOR FILING DATE: 1994-01-28  
 ; PRIOR APPLICATION NUMBER: US 08/318,519  
 ; PRIOR FILING DATE: 1994-10-05  
 ; NUMBER OF SEQ ID NOS: 10  
 ; SOFTWARE: Patentin version 3.0  
 ; SEQ ID NO 4  
 ; LENGTH: 151  
 ; TYPE: PRT  
 ; ORGANISM: Escherichia coli  
 US-09-741-873B-4

Query Match 66.5%; Score 520; DB 12; Length 151;  
 Best Local Similarity 68.2%; Pred. No. 2.2e-44;  
 Matches 103; Conservative 19; Mismatches 29; Indels 0; Gaps 0;  
 QY 1 MKLLKVAIAAIVFGSSALAGVVPQWGGGNGHGGGNSGDPSTLSIYQYGSANAALYDQ 60  
 Db 1 MKLLKVAIAAIVFGSSAVGVVPOYGGGNGHGGGNSGPNSELNIYQYGGNSALALQ 60  
 QY 61 LVTRVVTTHMAHAGYNGADVGGADNNTIELTQNGFNATIDOWNAKNSDITVQYGG 120  
 Db 61 TDARNSDLTITQHGCGGNGADVGGGDDSSIDLTORFGNSATLTDQWNGKNSMTVKQFGG 120  
 QY 121 NNAALVNQTSADSSVVMVQVGFNNATANQY 151  
 Db 121 GNGAAVDQTASNSSVNVTVQVGFNNATAHQY 151

RESULT 3  
 US-09-741-873B-2  
 ; Sequence 2, Application US/09741873B  
 ; Publication No. US20020081722A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Normark, Staffan  
 ; APPLICANT: Olsen, Arne  
 ; TITLE OF INVENTION: Fibronectin Binding Protein As Well As Its Preparation  
 ; FILE REFERENCE: 012889-084  
 ; CURRENT APPLICATION NUMBER: US/09/741,873B  
 ; CURRENT FILING DATE: 2003-04-04

; PRIOR APPLICATION NUMBER: SE 8801723-1  
 ; PRIOR FILING DATE: 1998-05-06  
 ; PRIOR APPLICATION NUMBER: US 08/978,878  
 ; PRIOR FILING DATE: 1997-11-26  
 ; PRIOR APPLICATION NUMBER: US 07/347,189  
 ; PRIOR FILING DATE: 1989-05-04  
 ; PRIOR APPLICATION NUMBER: US 07/789,437  
 ; PRIOR FILING DATE: 1991-11-06  
 ; PRIOR APPLICATION NUMBER: US 07/970,846  
 ; PRIOR FILING DATE: 1992-11-03  
 ; PRIOR APPLICATION NUMBER: US 08/187,865  
 ; PRIOR FILING DATE: 1994-01-28  
 ; PRIOR APPLICATION NUMBER: US 08/318,519  
 ; PRIOR FILING DATE: 1994-10-05  
 ; NUMBER OF SEQ ID NOS: 10  
 ; SOFTWARE: Patentin version 3.0  
 ; SEQ ID NO 2  
 ; LENGTH: 131  
 ; TYPE: PRT  
 ; ORGANISM: Escherichia coli  
 US-09-741-873B-2

Query Match 56.5%; Score 442; DB 12; Length 131;  
 Best Local Similarity 64.9%; Pred. No. 1.3e-36;  
 Matches 85; Conservative 18; Mismatches 28; Indels 0; Gaps 0;  
 QY 21 GVPVQWGGGNGHGGGNSGDPSTLSIYQYGSANAALYDQVTRVVTTHMAHAGYNGAD 80  
 Db 1 GVPVQYGGGNGHGGGNSGPNSELNIYQYGGNSALALQTDARNSDLTITQHGCGGNGAD 60  
 QY 81 VGGADNNTIELTQNGFNATIDOWNAKNSDITVQYGSANAALVNQTSADSSVVMVQV 140  
 Db 61 VGGSDSSIDLTORFGNSATLTDQWNGKNSMTVKQFGGNGAAVDQTASNSSVNVTVQV 120  
 QY 141 GFGNNATANQY 151  
 Db 121 GFGNNATAHQY 131

RESULT 4  
 US-09-741-873B-2  
 ; Sequence 2, Application US/09741873B  
 ; Publication No. US20040096965A9  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Normark, Staffan  
 ; APPLICANT: Olsen, Arne  
 ; TITLE OF INVENTION: Fibronectin Binding Protein As Well As Its Preparation  
 ; FILE REFERENCE: 012889-084  
 ; CURRENT APPLICATION NUMBER: US/09/741,873B  
 ; CURRENT FILING DATE: 2003-04-04  
 ; PRIOR APPLICATION NUMBER: SE 8801723-1  
 ; PRIOR FILING DATE: 1998-05-06  
 ; PRIOR APPLICATION NUMBER: US 08/978,878  
 ; PRIOR FILING DATE: 1997-11-26  
 ; PRIOR APPLICATION NUMBER: US 07/347,189  
 ; PRIOR FILING DATE: 1989-05-04  
 ; PRIOR APPLICATION NUMBER: US 07/789,437  
 ; PRIOR FILING DATE: 1991-11-06  
 ; PRIOR APPLICATION NUMBER: US 07/970,846  
 ; PRIOR FILING DATE: 1992-11-03  
 ; PRIOR APPLICATION NUMBER: US 08/187,865  
 ; PRIOR FILING DATE: 1994-01-28  
 ; PRIOR APPLICATION NUMBER: US 08/318,519  
 ; PRIOR FILING DATE: 1994-10-05  
 ; NUMBER OF SEQ ID NOS: 10  
 ; SOFTWARE: Patentin version 3.0  
 ; SEQ ID NO 2  
 ; LENGTH: 131  
 ; TYPE: PRT  
 ; ORGANISM: Escherichia coli  
 US-09-741-873B-2  
 Query Match 56.5%; Score 442; DB 12; Length 131;

Best Local Similarity 64.9%; Pred. No. 1.3e-36;  
Matches 85; Conservative 18; Mismatches 28; Indels 0; Gaps 0;  
QY 21 GVVPQGGGNGGSGSPDSTLSIYQGSANAALYDQLVTVVTHEMAHAGYGNAG 80  
DB 1 GVVPYGGGNGGSGNSGSELTIVYGGNSALALQDARNSDLTITQHGCGNGAD 60  
QY 81 VQGGADNSTIELTQNGFRNNATIDQWNAKSDITVGOYGGNNAALVNQOTASDSSVMVRQV 140  
DB 61 VQGGSDSSIDLQGFNSATLQWNGKNSMTVKQFGGNGAAVQDQASNSVNTQV 120  
QY 141 GFGNNATHQY 151  
DB 121 GFGNNATHQY 131  
RESULT 5  
US-10-369-493-20638  
; Sequence 20638, Application US/10369493  
; Publication No. US20030233675A1  
; GENERAL INFORMATION:  
; APPLICANT: Cao, Yongwei  
; APPLICANT: Hinkle, Gregory J.  
; APPLICANT: Slater, Steven C.  
; APPLICANT: Goldman, Barry S.  
; APPLICANT: Chen, Xianfeng  
; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF  
; FILE REFERENCE: 38-10(52052)B  
; CURRENT APPLICATION NUMBER: US/10/369,493  
; PRIOR FILING DATE: 2003-02-28  
; PRIOR APPLICATION NUMBER: US 60/360,039  
; PRIOR FILING DATE: 2002-02-21  
; NUMBER OF SEQ ID NOS: 47374  
; SEQ ID NO 20638  
; LENGTH: 445  
; TYPE: PRT  
; ORGANISM: Rhodospseudomonas palustris  
; FEATURE:  
; NAME/KEY: unsure  
; LOCATION: (1)..(445)  
; OTHER INFORMATION: unsure at all xaa locations  
US-10-369-493-20638  
Query Match 14.9%; Score 116.5; DB 15; Length 445;  
Best Local Similarity 27.5%; Pred. No. 0.0032;  
Matches 46; Conservative 20; Mismatches 60; Indels 41; Gaps 6;  
QY 7 AAFAA-----IWSGSALAGVVPQWGGG-----NHNGG-----GNSGSPDSTLSIYQY 50  
DB 19 AAFAADSNVTLNQTGNDQQAQITQSGNGSVGAENGSGFLQENGTLGSA-NLLTVKQS 77  
QY 51 GSANAALYDQLVTVVTHEMAHAGYGNAGDVCGGADNSTIELTQNGFRNNATIDQWNAK 110  
DB 78 GNSNSVGRD-----IQGKSGAGNSAAIFQEGTGSDELQQTGTSGAVPSSGWNWTN 129  
QY 111 -----SDITVGOYGGNNAALVNQOTASDSSVMVRQV 141  
DB 130 DPGVFNKITDSSNGSKSVIQDGKKNVFSIKQNTGNTSVMQIG 176  
RESULT 6  
US-09-793-306-146  
; Sequence 146, Application US/09793306  
; Patent No. US20020098200A1  
; GENERAL INFORMATION:  
; APPLICANT: Campos-Neto, Antonio  
; APPLICANT: Skeiky, Yasir  
; APPLICANT: Owendale, Pamela  
; APPLICANT: Jen, Shyian  
; APPLICANT: Lodes, Michael  
; APPLICANT: Corixa Corporation  
; TITLE OF INVENTION: Compounds and Methods for Diagnosis and Immunotherapy

TITLE OF INVENTION: of Tuberculosis  
FILE REFERENCE: 014058-008740US  
CURRENT APPLICATION NUMBER: US/09/793,306  
CURRENT FILING DATE: 2001-02-26  
PRIOR APPLICATION NUMBER: US 60/185,037  
PRIOR FILING DATE: 2000-02-25  
PRIOR APPLICATION NUMBER: US 60/223,828  
PRIOR FILING DATE: 2000-08-08  
NUMBER OF SEQ ID NOS: 164  
SOFTWARE: PatentIn Ver. 2.1  
SEQ ID NO 146  
LENGTH: 597  
TYPE: PRT  
ORGANISM: Artificial Sequence  
FEATURE:  
OTHER INFORMATION: Description of Artificial Sequence: mTCH3-His  
US-09-793-306-146  
Query Match 13.4%; Score 105; DB 9; Length 597;  
Best Local Similarity 29.0%; Pred. No. 0.065;  
Matches 40; Conservative 18; Mismatches 58; Indels 22; Gaps 7;  
QY 14 VGSALAGVVPQWGGGNGGSGSPDSTLSIYQGSANAALYDQLVTVVTHEMAHA 73  
DB 338 LTGDNLVGICALNSGIGN-LGFGNSG--NNNIGFNSGNNNVGFNS-----GNNNF 386  
QY 74 GYNGADVCGGADNSTIELTQNGFRNNATIDQ--WNAKSDITVGOYGGNNAALVNQOTAS 131  
DB 387 GFGNAGDINTGFGNAGD--TNTGFGNAGFNGMIGNAGNEDMGVNGSGFNVGVGN--AG 442  
QY 132 DSSVMVRQVGFNNATAN 149  
DB 443 NQS-----VGFGNAGTLN 455  
RESULT 7  
US-09-996-634-131  
; Sequence 131, Application US/09996634  
; Patent No. US20020172684A1  
; GENERAL INFORMATION:  
; APPLICANT: Nano, Francis  
; TITLE OF INVENTION: Mycobacterium Tuberculosis DNA Sequences Encoding  
; FILE REFERENCE: 61260  
; CURRENT APPLICATION NUMBER: US/09/996,634  
; CURRENT FILING DATE: 2001-11-28  
; PRIOR APPLICATION NUMBER: 09/447,135  
; PRIOR FILING DATE: 2000-01-03  
; PRIOR APPLICATION NUMBER: 09/990,823  
; PRIOR FILING DATE: 1997-12-15  
; PRIOR APPLICATION NUMBER: US 96/10375  
; PRIOR FILING DATE: 1996-06-14  
; PRIOR APPLICATION NUMBER: 60/000,254  
; PRIOR FILING DATE: 1995-06-15  
; NUMBER OF SEQ ID NOS: 169  
SOFTWARE: PatentIn Ver. 2.0  
SEQ ID NO 131  
LENGTH: 943  
TYPE: PRT  
ORGANISM: Mycobacterium tuberculosis  
US-09-996-634-131  
Query Match 12.8%; Score 100; DB 9; Length 943;  
Best Local Similarity 26.0%; Pred. No. 0.37;  
Matches 40; Conservative 12; Mismatches 52; Indels 50; Gaps 7;  
QY 16 GSALAGVVPQWGGG-GNHN-GGNSGSPDSTLSIYQGSANAALYDQLVTVVTHEMAHA 73  
DB 571 GSGNIGFNVGSGSLGNVNTGSGN-----LGYNIGFNVGVGY-----NV 610  
QY 74 GYNGADVCGGADNSTIELTQNGFRNNATIDQWNAKSDITVGOYGN----- 121  
DB 611 GFGNAGDFNQGFANT-----GNNNIGFANTGNNNIGLSDGNGQGFNTASGWN 660

QY	122	---	NAALVYNQTASDSSVM---VROVGFGNNATAN	149
:	:	:	:	:
DB	661	GTCNSGLFNGSTGNVGVFNAGTGNVGIANSCTGN	694	
 RESULT 8				
US-09-997-182-131				
; Sequence 131, Application US/09997182				
; Publication NO. US20030049269A1				
; GENERAL INFORMATION:				
; APPLICANT: Nano, Francis				
; TITLE OF INVENTION: Mycobacterium Tuberculosis DNT				
; FILE REFERENCE: 61258				
; CURRENT APPLICATION NUMBER: US/09/997,182				
; CURRENT FILING DATE: 2001-11-28				
; PRIOR APPLICATION NUMBER: 09/447,135				
; PRIOR FILING DATE: 2000-01-03				
; PRIOR APPLICATION NUMBER: 08/990,823				
; PRIOR FILING DATE: 1997-12-15				
; PRIOR APPLICATION NUMBER: US 96/10375				
; PRIOR FILING DATE: 1996-06-14				
; PRIOR APPLICATION NUMBER: 60/000,254				
; PRIOR FILING DATE: 1995-06-15				
; NUMBER OF SEQ ID NOS: 169				
; SOFTWARE: PatentIn Ver. 2.0				
; SEQ ID NO 131				
; LENGTH: 943				
; TYPE: PRT				
; ORGANISM: Mycobacterium tuberculosis				
US-09-997-182-131				
 Query Match 12.8%; Score 100; DB 10;				
Best Local Similarity 26.0%; Pred. No. 0.37; Mismatches 40; Conservative 12;				
QY	16	GSALAGVVPWGCGG-GNHN-CGGNSSGPDSTLGYQGVSIGDNI	549	
:	:	:	:	
DB	571	GSNGIGVFNVGSGSLGNVNI	581	
QY	74	GYGNGADYVCGADNSTIELFTQGFERNATIDQNANKNSI	549	
:	:	:	:	
DB	611	GFEGNAGDFNQGFANT-----GNNNICFANTGNNT	594	
QY	122	---	NAALVYNQTASDSSVM---VROVGFGNNATAN	149
:	:	:	:	:
DB	661	GTCNSGLFNGSTGNVGVFNAGTGNVGIANSCTGN	694	
 RESULT 9				
US-09-997-181-131				
; Sequence 131, Application US/09997181				
; Publication NO. US20030049269A1				
; GENERAL INFORMATION:				
; APPLICANT: Nano, Francis				
; TITLE OF INVENTION: Mycobacterium Tuberculosis DNT				
; FILE REFERENCE: 61257				
; CURRENT APPLICATION NUMBER: US/09/997,181				
; CURRENT FILING DATE: 2001-11-28				
; PRIOR APPLICATION NUMBER: 09/447,135				
; PRIOR FILING DATE: 2000-01-03				
; PRIOR APPLICATION NUMBER: 08/990,823				
; PRIOR FILING DATE: 1997-12-15				
; PRIOR APPLICATION NUMBER: US 96/10375				
; PRIOR FILING DATE: 1996-06-14				
; PRIOR APPLICATION NUMBER: 60/000,254				
; PRIOR FILING DATE: 1995-06-15				
; NUMBER OF SEQ ID NOS: 169				
; SOFTWARE: PatentIn Ver. 2.0				
; SEQ ID NO 131				
; LENGTH: 943				

Qy	16	GSALAGVVPWGGG-GNHN-GGNSSGPDSITLIYQGSANAALYDQLVTRVVTHEMAHA	73
Db	464	GSNGIGVFNVGSGSLGNVIGSGN-----LGIYNI GFGNVGDY-----	NV 503
Qy	74	GYGNGADVQGGADNSTIELTQNGFRNNTIDQWNAKNSDITVGOYGGN-----	121
Db	504	GFGNAGDFNQGFANT-----GNNIGFANTGNNIGIGLSGDNQOQGFNIASGWN	553
Qy	122	---NAALVNQTASDSSVM---VRQVFGNNATAN	149
Db	554	GTGNSGLFNSGTNNVGIFNAGTGNVGIANSGTGN	587
RESULT 12			
US-10-282-122A-64369			
; Sequence 64369, Application US/10282122A			
; Publication No. US20040029129A1			
; GENERAL INFORMATION:			
; APPLICANT: Wang, Liangsu			
; APPLICANT: Zamudio, Carlos			
; APPLICANT: Malone, Cheryl			
; APPLICANT: Haselbeck, Robert			
; APPLICANT: Ohsen, Kari			
; APPLICANT: Zyskind, Judith			
; APPLICANT: Wall, Daniel			
; APPLICANT: Trawick, John			
; APPLICANT: Carr, Grant			
; APPLICANT: Yamamoto, Robert			
; APPLICANT: Forsyth, R.			
; APPLICANT: Xu, H.			
; TITLE OF INVENTION: Identification of Essential Genes in Microorganisms			
; FILE REFERENCE: ELITRA.034A			
; CURRENT APPLICATION NUMBER: US/10/282,122A			
; CURRENT FILING DATE: 2003-02-20			
; PRIOR APPLICATION NUMBER: 60/191,078			
; PRIOR FILING DATE: 2000-03-21			
; PRIOR APPLICATION NUMBER: 60/206,848			
; PRIOR FILING DATE: 2000-05-23			
; PRIOR APPLICATION NUMBER: 60/207,727			
; PRIOR FILING DATE: 2000-05-26			
; PRIOR APPLICATION NUMBER: 60/230,335			
; PRIOR FILING DATE: 2000-09-06			
; PRIOR APPLICATION NUMBER: 60/230,347			
; PRIOR FILING DATE: 2000-09-09			
; PRIOR APPLICATION NUMBER: 60/242,578			
; PRIOR FILING DATE: 2000-10-23			
; PRIOR APPLICATION NUMBER: 60/253,625			
; PRIOR FILING DATE: 2000-11-27			
; PRIOR APPLICATION NUMBER: 60/257,931			
; PRIOR FILING DATE: 2000-12-22			
; PRIOR APPLICATION NUMBER: 60/267,636			
; PRIOR FILING DATE: 2001-02-09			
; PRIOR APPLICATION NUMBER: 60/269,308			
; PRIOR FILING DATE: 2001-02-16			
; Remaining Prior Application data removed - See File Wrapper or PALM.			
; NUMBER OF SEQ ID NOS: 78614			
; SOFTWARE: Patent in version 3.1			
; SEQ ID NO 64369			
; LENGTH: 3300			
; TYPE: PRT			
; ORGANISM: Mycobacterium tuberculosis			
US-10-282-122A-64369			
Query Match 12.8%; Score 100; DB 12; Length 3300;			
Best Local Similarity 26.0%; Pred. No. 1.7;			
Matches 40; Conservative 12; Mismatches 52; Indels 50; Gaps 7;			
Qy	16	GSALAGVVPWGGG-GNHN-GGNSSGPDSITLIYQGSANAALYDQLVTRVVTHEMAHA	73
Db	590	GSNGIGVFNVGSGSLGNVIGSGN-----LGIYNI GFGNVGDY-----	NV 629
Qy	74	GYGNGADVQGGADNSTIELTQNGFRNNTIDQWNAKNSDITVGOYGGN-----	121

Qy	16	GSALAGVVPWGGG-GNHN-GGNSSGPDSITLIYQGSANAALYDQLVTRVVTHEMAHA	73
Db	464	GSNGIGVFNVGSGSLGNVIGSGN-----LGIYNI GFGNVGDY-----	NV 503
Qy	74	GYGNGADVQGGADNSTIELTQNGFRNNTIDQWNAKNSDITVGOYGGN-----	121
Db	504	GFGNAGDFNQGFANT-----GNNIGFANTGNNIGIGLSGDNQOQGFNIASGWN	553
Qy	122	---NAALVNQTASDSSVM---VRQVFGNNATAN	149
Db	554	GTGNSGLFNSGTNNVGIFNAGTGNVGIANSGTGN	587
RESULT 11			
US-10-084-843-204			
; Sequence 204, Application US/10084843			
; Publication No. US20030143243A1			
; GENERAL INFORMATION:			
; APPLICANT: Reed, Steven G.			
; APPLICANT: Skeiky, Yasir A.W.			
; APPLICANT: Dillon, Davin C.			
; APPLICANT: Campos-Neto, Antonio			
; APPLICANT: Houghton, Raymond			
; APPLICANT: Vecvick, Thomas S.			
; APPLICANT: Twardzik, Daniel R.			
; APPLICANT: Lodes, Michael J.			
; APPLICANT: Hendrickson, Ronald C.			
; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR IMMUNOTHERAPY			
; TITLE OF INVENTION: AND DIAGNOSIS OF TUBERCULOSIS			
; NUMBER OF SEQUENCES: 355			
; CORRESPONDENCE ADDRESS:			
; ADDRESSEE: SEED AND BERRY LLP			
; STREET: 6300 Columbia Center, 701 Fifth Avenue			
; CITY: Seattle			
; STATE: Washington			
; COUNTRY: USA			
; ZIP: 98104-7092			
; COMPUTER READABLE FORM:			
; MEDIUM TYPE: Floppy disk			
; COMPUTER: IBM PC compatible			
; OPERATING SYSTEM: PC-DOS/MS-DOS			
; SOFTWARE: Patent in Release #1.0, Version #1.30			
; CURRENT APPLICATION DATA: US/10/084,843			
; APPLICATION NUMBER: US/10/084,843			
; FILING DATE: 25-Feb-2002			
; CLASSIFICATION: <Unknown>			
; PRIOR APPLICATION DATA:			
; APPLICATION NUMBER: US/09/072,967			
; FILING DATE: 05-MAY-1998			
; ATTORNEY/AGENT INFORMATION:			
; NAME: Maki, David J.			
; REGISTRATION NUMBER: 31,392			
; REFERENCE/DOCKET NUMBER: 210121.411C9			
; TELECOMMUNICATION INFORMATION:			
; TELEPHONE: (206) 622-4900			
; TELEFAX: (206) 682-6031			
; INFORMATION FOR SEQ ID NO: 204:			
; SEQUENCE CHARACTERISTICS:			
; LENGTH: 943 amino acids			
; TYPE: amino acid			
; STRANDEDNESS: <Unknown>			
; TOPOLOGY: linear			
; SEQUENCE DESCRIPTION: SEQ ID NO: 204:			
US-10-084-843-204			
Query Match 12.8%; Score 100; DB 14; Length 943;			
Best Local Similarity 26.0%; Pred. No. 0.37;			
Matches 40; Conservative 12; Mismatches 52; Indels 50; Gaps 7;			

Db 630 GFCNAGDFNOGFANT-----GNNIGFANTGNNNIGLIGLSDNQNGFNASGWSN 679  
Qy 122 ---NAALVNQTSADSSVM---VRQVGFNGNATAN 149  
Db 680 GTGNSGLFNGTNNVNGIFNAGTNGVNGIANSNGTN 713

RESULT 13  
US-10-369-493-20619  
; Sequence 20619, Application US/10369493  
; Publication No. US20030233675A1  
; GENERAL INFORMATION:  
; APPLICANT: Cao, Yongwei  
; APPLICANT: Hinkle, Gregory J.  
; APPLICANT: Slater, Steven C.  
; APPLICANT: Goldman, Barry S.  
; APPLICANT: Chen, Xianfeng  
; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF  
; FILE REFERENCE: 38-10(52052)B  
; CURRENT APPLICATION NUMBER: US/10/369,493  
; PRIOR FILING DATE: 2003-02-28  
; PRIOR APPLICATION NUMBER: US 60/360,039  
; PRIOR FILING DATE: 2002-02-21  
; NUMBER OF SEQ ID NOS: 473/4  
; SEQ ID NO 20619  
; LENGTH: 486  
; TYPE: PRT  
; ORGANISM: Rhodospseudomonas palustris  
; FEATURE:  
; NAME/KEY: unsure  
; LOCATION: (1)..(486)  
; OTHER INFORMATION: unsure at all Xaa locations  
US-10-369-493-20619

Query Match 12.6%; Score 98.5; DB 15; Length 486;  
Best Local Similarity 25.7%; Pred. No. 0.23;  
Matches 37; Conservative 22; Mismatches 50; Indels 35; Gaps 6;  
Qy 11 AIVWGS---ALAGVVPQWGGGNGHNGGNSGP-----DSTLSIYQYGSANAA 57  
Db 86 SVVAGTDYRDIVAGVLPGLGKXAYAGDGGSPFAFFKDLASDTPFSAPEFGAAGF- 144  
Qy 58 YQLVTRVTVTHMAGYNGADVCGADNSTIELTQNGFRNATIDQWNAKNSDITVG- 116  
Db 145 -----VIHTAAGVRPRGALISQG--NLLI-----AQSSLVDARLITRQADNVLM 188

Qy 117 ----QYGGNNAALVNQTSADSSVM 136  
Db 189 LPLFHTVGLGLMLTLQAGGASVI 212

RESULT 14  
US-10-282-122A-49412  
; Sequence 49412, Application US/10282122A  
; Publication No. US20040029129A1  
; GENERAL INFORMATION:  
; APPLICANT: Wang, Liangsu  
; APPLICANT: Zamudio, Carlos  
; APPLICANT: Malone, Cheryl  
; APPLICANT: Haselbeck, Robert  
; APPLICANT: Ohlsen, Kari  
; APPLICANT: Zyskind, Judith  
; APPLICANT: Wall, Daniel  
; APPLICANT: Trawick, John  
; APPLICANT: Carx, Grant  
; APPLICANT: Yamamoto, Robert  
; APPLICANT: Forsyth, R.  
; APPLICANT: Xu, H.  
; TITLE OF INVENTION: Identification of Essential Genes in Microorganisms  
; FILE REFERENCE: ELITRA.034A  
; CURRENT APPLICATION NUMBER: US/10/282,122A  
; CURRENT FILING DATE: 2003-02-20

; PRIOR APPLICATION NUMBER: 60/191,078  
; PRIOR FILING DATE: 2000-03-21  
; PRIOR APPLICATION NUMBER: 60/206,848  
; PRIOR FILING DATE: 2000-05-23  
; PRIOR APPLICATION NUMBER: 60/207,727  
; PRIOR FILING DATE: 2000-05-26  
; PRIOR APPLICATION NUMBER: 60/230,335  
; PRIOR FILING DATE: 2000-09-06  
; PRIOR APPLICATION NUMBER: 60/230,347  
; PRIOR FILING DATE: 2000-09-09  
; PRIOR APPLICATION NUMBER: 60/242,578  
; PRIOR FILING DATE: 2000-10-23  
; PRIOR APPLICATION NUMBER: 60/253,625  
; PRIOR FILING DATE: 2000-11-27  
; PRIOR APPLICATION NUMBER: 60/257,931  
; PRIOR FILING DATE: 2000-12-22  
; PRIOR APPLICATION NUMBER: 60/267,636  
; PRIOR FILING DATE: 2001-02-09  
; PRIOR APPLICATION NUMBER: 60/269,308  
; PRIOR FILING DATE: 2001-02-16  
; Remaining Prior Application data removed - See File Wrapper or PALM.  
; NUMBER OF SEQ ID NOS: 78614  
; SOFTWARE: PatentIn version 3.1  
; SEQ ID NO 49412  
; LENGTH: 186  
; TYPE: PRT  
; ORGANISM: Burkholderia fungorum  
US-10-282-122A-49412

Query Match 12.3%; Score 96; DB 12; Length 186;  
Best Local Similarity 29.3%; Pred. No. 0.12;  
Matches 49; Conservative 18; Mismatches 62; Indels 38; Gaps 10;  
Qy 1 MKLLKVAFAAIVVSGSALAGVVPQWGGGNGHNGGNS-----GPDSTLSIYQYGSANAA 56  
Db 1 MKL-----SLPATAAASSLFLVPL-----GAHAADGTISITGTVDSTCSI--NGNANGT 50  
Qy 57 LYDQLVTRVTVTHMAGYNGADVG--QCADNST--IELTQNG-----FRNNAT 102  
Db 51 PADKAIT-----LAIVPAGSLASAGAVAGTSPDQLSLGCTGTATKATARFENGPT 104  
Qy 103 IDQWNAKNSDITVQYGGNNAALVNQTSADSSVMVRQVGFNGNATAN 149  
Db 105 VDQTNGVLSN-TAGTAQNVREVRLLNAQMOPINV---TTGANNDITTN 147

RESULT 15  
US-09-952-267-5  
; Sequence 5, Application US/09952267  
; Publication No. US20030032772A1  
; GENERAL INFORMATION:  
; APPLICANT: HANSEN, ERIC J.  
; APPLICANT: AEBI, CHRISTOPH  
; APPLICANT: COPE, LESLIE D.  
; APPLICANT: MACIVER, ISOBEL  
; APPLICANT: FISKE, MICHAEL J.  
; APPLICANT: FREDENBURG, ROSS A.  
; TITLE OF INVENTION: USP41 AND USP42 ANTIGENS OF MORAXELLA CATARRHALIS  
; FILE REFERENCE: AMCY.024  
; CURRENT APPLICATION NUMBER: US/09/952,267  
; CURRENT FILING DATE: 2001-09-12  
; PRIOR APPLICATION NUMBER: 09/336,447  
; PRIOR FILING DATE: 1999-06-21  
; NUMBER OF SEQ ID NOS: 98  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 5  
; LENGTH: 892  
; TYPE: PRT  
; ORGANISM: Moraxella catarrhalis  
US-09-952-267-5

Query Match 11.9%; Score 93; DB 10; Length 892;  
Best Local Similarity 27.7%; Pred. No. 1.7;

		Matches	41; Conservative	14; Mismatches	49; Indels	44; Gaps	9;
QY	28	GGGNHN-----	GGGNSS--	CPDSTLSIYQYGSANAALYDQLVTRVVTHEMAHAGY	75		
Db	89	GGGDYNEAKGNYSTVGGSSNTAKGEKSTIGGGDTNDANG-----	TYSTIGGGY	137			
QY	76	-----GNGADVGGADNSTI--ELTQNGFRNNATTIDQWNAKNSDITVQYG--	GNNAAAL	125			
Db	138	YSRAIGDSSTIGGGYINQATGEKSTVAGGRNN---QATGNNSTVAGGSYNQATGNNSTV	193				
QY	126	V-----NOTASDSSVMVRQVFGGNATAN	149				
Db	194	AGGSHNQATGEGSF---AAGVENKANAN	218				

Search completed: August 2, 2004, 15:36:13  
Job time : 37.8 secs

This Page Blank (uspto)



GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: August 2, 2004, 14:48:33 ; Search time 167.9 Seconds  
(without alignments)  
877.809 Million cell updates/sec

Title: US-09-543-407-26

Perfect score: 782

Sequence: 1 MKLLKVAFAAIVVSGSALA.....DSSVMVRQVGFNNATANQY 151

Scoring table:

BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 6019581 seqs, 976053577 residues

Total number of hits satisfying chosen parameters: 6019581

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

Pending Patents AA Main: \*  
1: /cgn2\_6/ptodata/2/paa/PCTUS COMB.pcp.\*  
2: /cgn2\_6/ptodata/2/paa/US06 COMB.pcp.\*  
3: /cgn2\_6/ptodata/2/paa/US07 COMB.pcp.\*  
4: /cgn2\_6/ptodata/2/paa/US08 COMB.pcp.\*  
5: /cgn2\_6/ptodata/2/paa/US081 COMB.pcp.\*  
6: /cgn2\_6/ptodata/2/paa/US082 COMB.pcp.\*  
7: /cgn2\_6/ptodata/2/paa/US083 COMB.pcp.\*  
8: /cgn2\_6/ptodata/2/paa/US084 COMB.pcp.\*  
9: /cgn2\_6/ptodata/2/paa/US085 COMB.pcp.\*  
10: /cgn2\_6/ptodata/2/paa/US086 COMB.pcp.\*  
11: /cgn2\_6/ptodata/2/paa/US087 COMB.pcp.\*  
12: /cgn2\_6/ptodata/2/paa/US088 COMB.pcp.\*  
13: /cgn2\_6/ptodata/2/paa/US089 COMB.pcp.\*  
14: /cgn2\_6/ptodata/2/paa/US090 COMB.pcp.\*  
15: /cgn2\_6/ptodata/2/paa/US091 COMB.pcp.\*  
16: /cgn2\_6/ptodata/2/paa/US092 COMB.pcp.\*  
17: /cgn2\_6/ptodata/2/paa/US093 COMB.pcp.\*  
18: /cgn2\_6/ptodata/2/paa/US094 COMB.pcp.\*  
19: /cgn2\_6/ptodata/2/paa/US095 COMB.pcp.\*  
20: /cgn2\_6/ptodata/2/paa/US096 COMB.pcp.\*  
21: /cgn2\_6/ptodata/2/paa/US097A COMB.pcp.\*  
22: /cgn2\_6/ptodata/2/paa/US097B COMB.pcp.\*  
23: /cgn2\_6/ptodata/2/paa/US098 COMB.pcp.\*  
24: /cgn2\_6/ptodata/2/paa/US099A COMB.pcp.\*  
25: /cgn2\_6/ptodata/2/paa/US099B COMB.pcp.\*  
26: /cgn2\_6/ptodata/2/paa/US100 COMB.pcp.\*  
27: /cgn2\_6/ptodata/2/paa/US101 COMB.pcp.\*  
28: /cgn2\_6/ptodata/2/paa/US102 COMB.pcp.\*  
29: /cgn2\_6/ptodata/2/paa/US103 COMB.pcp.\*  
30: /cgn2\_6/ptodata/2/paa/US104 COMB.pcp.\*  
31: /cgn2\_6/ptodata/2/paa/US105 COMB.pcp.\*  
32: /cgn2\_6/ptodata/2/paa/US106 COMB.pcp.\*  
33: /cgn2\_6/ptodata/2/paa/US107 COMB.pcp.\*  
34: /cgn2\_6/ptodata/2/paa/US108 COMB.pcp.\*  
35: /cgn2\_6/ptodata/2/paa/US109 COMB.pcp.\*  
36: /cgn2\_6/ptodata/2/paa/US110 COMB.pcp.\*  
37: /cgn2\_6/ptodata/2/paa/US111 COMB.pcp.\*  
38: /cgn2\_6/ptodata/2/paa/US112 COMB.pcp.\*  
39: /cgn2\_6/ptodata/2/paa/US113 COMB.pcp.\*  
40: /cgn2\_6/ptodata/2/paa/US114 COMB.pcp.\*  
41: /cgn2\_6/ptodata/2/paa/US115 COMB.pcp.\*  
42: /cgn2\_6/ptodata/2/paa/US116 COMB.pcp.\*  
43: /cgn2\_6/ptodata/2/paa/US117 COMB.pcp.\*  
44: /cgn2\_6/ptodata/2/paa/US118 COMB.pcp.\*  
45: /cgn2\_6/ptodata/2/paa/US119 COMB.pcp.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match Length	DB ID	Description
------------	-------	--------------------	-------	-------------

1	782	100.0	151	19	US-09-543-407-26	Sequence 26, Appl
2	709	90.7	151	19	US-09-543-407-20	Sequence 20, Appl
3	692	88.5	151	19	US-09-543-407-5	Sequence 5, Appl
4	687	87.9	151	6	US-08-233-642A-57	Sequence 57, Appl
5	675	86.3	151	19	US-09-543-407-18	Sequence 18, Appl
6	617	78.9	151	19	US-09-543-407-28	Sequence 28, Appl
7	614	78.5	151	19	US-09-543-407-12	Sequence 12, Appl
8	612	78.3	151	19	US-09-543-407-14	Sequence 14, Appl
9	609	77.9	151	19	US-09-543-407-24	Sequence 24, Appl
10	605	77.4	151	19	US-09-543-407-31	Sequence 31, Appl
11	601	76.9	151	19	US-09-543-407-22	Sequence 22, Appl
12	600	76.7	151	19	US-09-543-407-30	Sequence 30, Appl
13	577	73.8	151	19	US-09-543-407-16	Sequence 16, Appl
14	523	66.9	151	19	US-09-543-407-7	Sequence 7, Appl
15	520	66.5	151	13	US-08-978-878-4	Sequence 4, Appl
16	520	66.5	151	21	US-09-741-873B-4	Sequence 4, Appl
17	518	66.2	151	33	US-60-352-948-2	Sequence 2, Appl
18	518	66.2	151	33	US-60-444-371-2	Sequence 2, Appl
19	507	64.8	120	6	US-08-233-642A-55	Sequence 55, Appl
20	473	60.5	109	19	US-09-543-407-34	Sequence 34, Appl
21	466	59.6	158	16	US-09-252-691-5834	Sequence 5834, Ap
22	466	59.6	158	16	US-09-252-691C-5834	Sequence 5834, Ap
23	466	59.6	158	30	US-10-417-886-5834	Sequence 5834, Ap
24	442	56.5	131	13	US-08-978-878-2	Sequence 2, Appl
25	442	56.5	131	21	US-09-741-873B-2	Sequence 2, Appl
26	338	43.2	109	19	US-09-543-407-35	Sequence 35, Appl
27	266	34.0	68	19	US-09-543-407-37	Sequence 37, Appl
28	201.5	25.8	70	19	US-09-543-407-32	Sequence 32, Appl
29	156	19.9	48	19	US-09-543-407-39	Sequence 39, Appl
30	116.5	14.9	186	16	US-09-252-691-5833	Sequence 5833, Ap
31	116.5	14.9	186	16	US-09-252-691C-5833	Sequence 5833, Ap
32	116.5	14.9	186	30	US-10-417-886-5833	Sequence 5833, Ap
33	116.5	14.9	445	29	US-10-369-493-20638	Sequence 20638, A
34	116.5	14.9	445	33	US-60-360-039-20638	Sequence 20638, A
35	114.5	14.6	145	21	US-09-739-449-8854	Sequence 8854, Ap
36	114.5	14.6	145	23	US-09-803-110-8854	Sequence 8854, Ap
37	112	14.3	151	19	US-09-543-407-8	Sequence 8, Appl
38	107	13.7	151	19	US-09-543-407-6	Sequence 6, Appl
39	105	13.4	597	1	PCT-US01-05992-146	Sequence 146, App
40	105	13.4	597	22	US-09-793-306-146	Sequence 146, App
41	100	12.8	943	1	PCT-US99-03265-199	Sequence 199, App
42	100	12.8	943	1	PCT-US99-03268-204	Sequence 204, App
43	100	12.8	943	13	US-08-942-341-199	Sequence 199, App
44	100	12.8	943	13	US-08-942-578-204	Sequence 204, App
45	100	12.8	943	14	US-09-024-753-199	Sequence 199, App

#### ALIGNMENTS

RESULT 1  
US-09-543-407-26  
; Sequence 26, Application US/09543407  
; GENERAL INFORMATION:  
; APPLICANT: White, Aaron P.  
; APPLICANT: Doran, James L.  
; APPLICANT: Collinson, S. Karen  
; APPLICANT: Kay, William W.  
; TITLE OF INVENTION: BACTERIAL FIMBRIAL SYSTEM FOR PRESENTATION OF HETEROLOGOUS PEPTIDE SEQUENCES  
; FILE REFERENCE: 920043.406  
; CURRENT APPLICATION NUMBER: US/09/543,407  
; CURRENT FILING DATE: 2000-04-05  
; NUMBER OF SEQ ID NOS: 59  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 26  
; LENGTH: 151  
; TYPE: PPT  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: Recombinant Salmonella enteritidis 3b afgA  
; OTHER INFORMATION: sequence containing the replacement fragment  
; OTHER INFORMATION: encoding PT3 from GP63 of Leishmania major.

```
US-09-543-407-26
Query Match      100.0%; Score 782; DB 19; Length 151;
Best Local Similarity 100.0%; Pred. No. 1.8e-75;
Matches 151; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MKLLKVAFAAIVVSGSALAGVVPQWGGGNGHNGGNSGDPSTLSIYQYGSANAALVDQ 60
   |||||
DB 1 MKLLKVAFAAIVVSGSALAGVVPQWGGGNGHNGGNSGDPSTLSIYQYGSANAALVDQ 60
   |||||

QY 61 LVTRVVTHEMAHAGYNGGADVGGGADNSTIETQNGFRNNATIDOWNAKNSDITVQYGG 120
   |||||
DB 61 LVTRVVTHEMAHAGYNGGADVGGGADNSTIETQNGFRNNATIDOWNAKNSDITVQYGG 120
   |||||

QY 121 NNAALVNQTASDSSVMVROVGFGNNTANQY 151
   |||||
DB 121 NNAALVNQTASDSSVMVROVGFGNNTANQY 151
   |||||

RESULT 2
US-09-543-407-20
; Sequence 20, Application US/09543407
; GENERAL INFORMATION:
; APPLICANT: White, Aaron P.
; APPLICANT: Doran, James L.
; APPLICANT: Collinson, S. Karen
; APPLICANT: Kay, William W.
; TITLE OF INVENTION: BACTERIAL FIMBRIAL SYSTEM FOR
; FILE REFERENCE: 920043.406
; CURRENT APPLICATION NUMBER: US/09/543,407
; CURRENT FILING DATE: 2000-04-05
; NUMBER OF SEQ ID NOS: 59
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 20
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Recombinant Salmonella enteritidis 3b afga
; OTHER INFORMATION: sequence containing the replacement fragment
; OTHER INFORMATION: encoding PT3 from GP63 of Leishmania major.
US-09-543-407-20

Query Match      90.7%; Score 709; DB 19; Length 151;
Best Local Similarity 89.9%; Pred. No. 1.3e-67;
Matches 143; Conservative 0; Mismatches 0; Indels 16; Gaps 2;

QY 1 MKLLKVAFAAIVVSGSALAGVVPQWGGGNGHNGGNSGDPSTLSIYQYGSANAAL--- 57
   |||||
DB 1 MKLLKVAFAAIVVSGSALAGVVPQWGGGNGHNGGNSGDPSTLSIYQYGSANAALQ 60
   |||||

QY 58 -----YDQLVTRVVTHEMAHAGYNGGADVGGGADNSTIETQNGFRNNATIDOWNAKNSD 112
   |||||
DB 61 SDARKYDQLVTRVVTHEMAFA-----GQADNSTIETQNGFRNNATIDOWNAKNSD 112
   |||||

QY 113 ITVQYGGNNAALVNQTASDSSVMVROVGFGNNTANQY 151
   |||||
DB 113 ITVQYGGNNAALVNQTASDSSVMVROVGFGNNTANQY 151
   |||||

RESULT 3
US-09-543-407-5
; Sequence 5, Application US/09543407
; GENERAL INFORMATION:
; APPLICANT: White, Aaron P.
; APPLICANT: Doran, James L.
; APPLICANT: Collinson, S. Karen
; APPLICANT: Kay, William W.
; TITLE OF INVENTION: BACTERIAL FIMBRIAL SYSTEM FOR
; FILE REFERENCE: 920043.406
; CURRENT APPLICATION NUMBER: US/09/543,407
```

```
; CURRENT FILING DATE: 2000-04-05
; NUMBER OF SEQ ID NOS: 59
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 5
; LENGTH: 151
; TYPE: PRT
; ORGANISM: Salmonella enteritidis
US-09-543-407-5

Query Match      88.5%; Score 692; DB 19; Length 151;
Best Local Similarity 90.7%; Pred. No. 8.7e-66;
Matches 137; Conservative 2; Mismatches 12; Indels 0; Gaps 0;

QY 1 MKLLKVAFAAIVVSGSALAGVVPQWGGGNGHNGGNSGDPSTLSIYQYGSANAALVDQ 60
   |||||
DB 1 MKLLKVAFAAIVVSGSALAGVVPQWGGGNGHNGGNSGDPSTLSIYQYGSANAALQ 60
   |||||

QY 61 LVTRVVTHEMAHAGYNGGADVGGGADNSTIETQNGFRNNATIDOWNAKNSDITVQYGG 120
   |||||
DB 61 SDARKSETTITQSGYNGGADVGGGADNSTIETQNGFRNNATIDOWNAKNSDITVQYGG 120
   |||||

QY 121 NNAALVNQTASDSSVMVROVGFGNNTANQY 151
   |||||
DB 121 NNAALVNQTASDSSVMVROVGFGNNTANQY 151
   |||||

RESULT 4
US-08-233-642A-57
; Sequence 57, Application US/08233642A
; GENERAL INFORMATION:
; APPLICANT: Kay, William W.
; APPLICANT: Collinson, S. Karen
; APPLICANT: Clouthier, Sharon C.
; APPLICANT: Doran, James L.
; TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR SALMONELLA-
; TITLE OF INVENTION: BASED VACCINES
; NUMBER OF SEQUENCES: 58
; NUMBER OF SEQUENCES: -
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Seed and Berry
; STREET: 6300 Columbia Center, 701 Fifth Avenue
; CITY: Seattle
; STATE: Washington
; COUNTRY: U.S.A.
; ZIP: 98104-7092
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/233,642A
; FILING DATE: 26-APR-1994
; CLASSIFICATION: 424
; ATTORNEY/AGENT INFORMATION:
; NAME: King, Joshua
; REGISTRATION NUMBER: 35,570
; REFERENCE/DOCKET NUMBER: 920043.403C3
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (206) 622-4900
; TELEFAX: (206) 682-6031
; TELEX: 3723836 SEEDANBERRY
; INFORMATION FOR SEQ ID NO: 57:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 151 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-233-642A-57

Query Match      87.9%; Score 687; DB 6; Length 151;
Best Local Similarity 90.1%; Pred. No. 3e-65;
Matches 136; Conservative 2; Mismatches 13; Indels 0; Gaps 0;
```

RESULT 8  
US-09-543-407-14  
; Sequence 14: Application US/09543407  
; GENERAL INFORMATION:  
; APPLICANT: White, Aaron P.  
; APPLICANT: Doran, James L.

APPLICANT: Collinson, S. Karen  
APPLICANT: Kay, William W.  
TITLE OF INVENTION: BACTERIAL FIMBRIAL SYSTEM FOR  
PRESENTATION OF HETEROLOGOUS PEPTIDE SEQUENCES  
FILE REFERENCE: 920043.406  
CURRENT APPLICATION NUMBER: US/09/543,407  
CURRENT FILING DATE: 2000-04-05  
NUMBER OF SEQ ID NOS: 59  
SOFTWARE: FastSeq for Windows Version 4.0  
SEQ ID NO 14  
LENGTH: 151  
TYPE: PRT  
ORGANISM: Artificial Sequence  
FEATURE:  
OTHER INFORMATION: Recombinant Salmonella enteritidis 3b afga  
OTHER INFORMATION: sequence containing the replacement fragment  
OTHER INFORMATION: encoding PT3 from GP63 of Leishmania major.  
US-09-543-407-14

Query Match 78.3%; Score 612; DB 19; Length 151;  
Best Local Similarity 81.5%; Pred. No. 3.6e-57;  
Matches 123; Conservative 5; Mismatches 23; Indels 0; Gaps 0;

Qy 1 MKLLKVAFAAIVVSGSALAGVVPQWGGGNGGNGSGGPDSTLSIYQYGSANAALYDQ 60  
Db 1 MKLLKVAFAAIVVSGSALAGVVPQWGGGNGGNGSGGPDSTLSIYQYGSANAALQ 60  
Qy 61 LVTRVVTHEMAHAGYNGADVQGGADNSTIETQNGFRNNATIDOWNAKNSDITVGYGG 120  
Db 61 SDARKSETTITQSGYNGADVQGGADNSTIETQNGFRNNATIDOWNAKNSDITVGYDQ 120  
Qy 121 NNAALVNQATSDSSVMVROVGFNNATANQY 151  
Db 121 LVTRVVTHEMAHAGVVMVROVGFNNATANQY 151

## RESULT 9

US-09-543-407-24

Sequence 24, Application US/09543407  
GENERAL INFORMATION:  
APPLICANT: White, Aaron P.  
APPLICANT: Doran, James L.  
APPLICANT: Collinson, S. Karen  
APPLICANT: Kay, William W.  
TITLE OF INVENTION: BACTERIAL FIMBRIAL SYSTEM FOR  
PRESENTATION OF HETEROLOGOUS PEPTIDE SEQUENCES  
FILE REFERENCE: 920043.406  
CURRENT APPLICATION NUMBER: US/09/543,407  
CURRENT FILING DATE: 2000-04-05  
NUMBER OF SEQ ID NOS: 59  
SOFTWARE: FastSeq for Windows Version 4.0  
SEQ ID NO 24  
LENGTH: 151  
TYPE: PRT  
ORGANISM: Artificial Sequence  
FEATURE:  
OTHER INFORMATION: Recombinant Salmonella enteritidis 3b afga  
OTHER INFORMATION: sequence containing the replacement fragment  
OTHER INFORMATION: encoding PT3 from GP63 of Leishmania major.

US-09-543-407-24

Query Match 77.9%; Score 609; DB 19; Length 151;  
Best Local Similarity 82.1%; Pred. No. 7.5e-57;  
Matches 124; Conservative 4; Mismatches 23; Indels 0; Gaps 0;

Qy 1 MKLLKVAFAAIVVSGSALAGVVPQWGGGNGGNGSGGPDSTLSIYQYGSANAALYDQ 60  
Db 1 MKLLKVAFAAIVVSGSALAGVVPQWGGGNGGNGSGGPDSTLSIYQYGSANAALQ 60  
Qy 61 LVTRVVTHEMAHAGYNGADVQGGADNSTIETQNGFRNNATIDOWNAKNSDITVGYGG 120  
Db 61 SDARKSETTITQSGYNGADVQGGADNSTIETQNGFRNNATIDOWNAKNYDQLVTRVT 120

Qy 121 NNAALVNQATSDSSVMVROVGFNNATANQY 151  
Db 121 HEMAHANQATSDSSVMVROVGFNNATANQY 151

## RESULT 10

US-09-543-407-31

Sequence 31, Application US/09543407  
GENERAL INFORMATION:  
APPLICANT: White, Aaron P.  
APPLICANT: Doran, James L.  
APPLICANT: Collinson, S. Karen  
APPLICANT: Kay, William W.  
TITLE OF INVENTION: BACTERIAL FIMBRIAL SYSTEM FOR  
PRESENTATION OF HETEROLOGOUS PEPTIDE SEQUENCES  
FILE REFERENCE: 920043.406  
CURRENT APPLICATION NUMBER: US/09/543,407  
CURRENT FILING DATE: 2000-04-05  
NUMBER OF SEQ ID NOS: 59  
SOFTWARE: FastSeq for Windows Version 4.0  
SEQ ID NO 31  
LENGTH: 131  
TYPE: PRT  
ORGANISM: Salmonella enteritidis  
US-09-543-407-31

Query Match 77.4%; Score 605; DB 19; Length 131;  
Best Local Similarity 89.3%; Pred. No. 1.7e-56;  
Matches 117; Conservative 2; Mismatches 12; Indels 0; Gaps 0;

Qy 21 GVVPQWGGGNGGNGSGGPDSTLSIYQYGSANAALYDQLVTRVVTHEMAHAGYNGAD 80  
Db 1 GVVPQWGGGNGGNGSGGPDSTLSIYQYGSANAALQSDARKSETTITQSGYNGAD 60  
Qy 81 VQGGADNSTIETQNGFRNNATIDOWNAKNSDITVGYGNNALVNQATSDSSVMVROV 140  
Db 61 VQGGADNSTIETQNGFRNNATIDOWNAKNSDITVGYGNNALVNQATSDSSVMVROV 120  
Qy 141 GFGNNATANQY 151  
Db 121 GFGNNATANQY 131

## RESULT 11

US-09-543-407-22

Sequence 22, Application US/09543407  
GENERAL INFORMATION:  
APPLICANT: White, Aaron P.  
APPLICANT: Doran, James L.  
APPLICANT: Collinson, S. Karen  
APPLICANT: Kay, William W.  
TITLE OF INVENTION: BACTERIAL FIMBRIAL SYSTEM FOR  
PRESENTATION OF HETEROLOGOUS PEPTIDE SEQUENCES  
FILE REFERENCE: 920043.406  
CURRENT APPLICATION NUMBER: US/09/543,407  
CURRENT FILING DATE: 2000-04-05  
NUMBER OF SEQ ID NOS: 59  
SOFTWARE: FastSeq for Windows Version 4.0  
SEQ ID NO 22  
LENGTH: 151  
TYPE: PRT  
ORGANISM: Artificial Sequence  
FEATURE:  
OTHER INFORMATION: Recombinant Salmonella enteritidis 3b afga  
OTHER INFORMATION: sequence containing the replacement fragment  
OTHER INFORMATION: encoding PT3 from GP63 of Leishmania major.

US-09-543-407-22

Query Match 76.9%; Score 601; DB 19; Length 151;  
Best Local Similarity 81.5%; Pred. No. 5.5e-56;  
Matches 123; Conservative 4; Mismatches 24; Indels 0; Gaps 0;

Qy 1 MKLLKVAFAAIVVSGSALAGVVPQWGGGNGGNGSGGPDSTLSIYQYGSANAALYDQ 60

Db 1 MKLLKVAFAAIVVSGSALAGVVPQWGGGNGHNGSSGPDSTLSIYQGSANAALQ 60  
QY 61 LVTRVVTHEMAHAGYNGADVGQADNSTIETQNGFRNNATIDQWNAKNSDITVGYGG 120  
Db 61 SDARKSETTITQSGYNGADVGQADNDYDQLVTRVVTHEMAHADQWNAKNSDITVGYGG 120  
QY 121 NNAALVNQTASDSSVMVRQVGFNNATANQY 151  
Db 121 NNAALVNQTASDSSVMVRQVGFNNATANQY 151

## RESULT 12

US-09-543-407-30  
; Sequence 30, Application US/09543407  
; GENERAL INFORMATION:  
; APPLICANT: White, Aaron P.  
; APPLICANT: Doran, James L.  
; APPLICANT: Collinson, S. Karen  
; APPLICANT: Kay, William W.  
; TITLE OF INVENTION: BACTERIAL FIMBRIAL SYSTEM FOR  
; TITLE OF INVENTION: PRESENTATION OF HETEROLOGOUS PEPTIDE SEQUENCES  
; FILE REFERENCE: 920043.406  
; CURRENT APPLICATION NUMBER: US/09/543,407  
; NUMBER OF SEQ ID NOS: 59  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 30  
; LENGTH: 151  
; TYPE: PRT  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: Recombinant Salmonella enteritidis 3b afgA  
; OTHER INFORMATION: sequence containing the replacement fragment  
; OTHER INFORMATION: encoding PT3 from GP63 of Leishmania major.  
US-09-543-407-30

Query Match 76.7%; Score 600; DB 19; Length 151;  
Best Local Similarity 81.5%; Pred. No. 7e-56;  
Matches 123; Conservative 4; Mismatches 24; Indels 0; Gaps 0;  
QY 1 MKLLKVAFAAIVVSGSALAGVVPQWGGGNGHNGSSGPDSTLSIYQGSANAALQ 60  
Db 1 MKLLKVAFAAIVVSGSALAGVVPQWGGGNGHNGSSGPDSTLSIYQGSANAALQ 60  
QY 61 LVTRVVTHEMAHAGYNGADVGQADNSTIETQNGFRNNATIDQWNAKNSDITVGYGG 120  
Db 61 SDARKSETTITQSGYNGADVGQADNSTIETQNGFRNNATIDQWNAKNSDITVGYGG 120  
QY 121 NNAALVNQTASDSSVMVRQVGFNNATANQY 151  
Db 121 NNAALVNQTASDSSVMVRQVGFNNATANQY 151

## RESULT 13

US-09-543-407-16  
; Sequence 16, Application US/09543407  
; GENERAL INFORMATION:  
; APPLICANT: White, Aaron P.  
; APPLICANT: Doran, James L.  
; APPLICANT: Collinson, S. Karen  
; APPLICANT: Kay, William W.  
; TITLE OF INVENTION: BACTERIAL FIMBRIAL SYSTEM FOR  
; TITLE OF INVENTION: PRESENTATION OF HETEROLOGOUS PEPTIDE SEQUENCES  
; FILE REFERENCE: 920043.406  
; CURRENT APPLICATION NUMBER: US/09/543,407  
; NUMBER OF SEQ ID NOS: 59  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 16  
; LENGTH: 151  
; TYPE: PRT  
; ORGANISM: Artificial Sequence

; FEATURE:  
; OTHER INFORMATION: Recombinant Salmonella enteritidis 3b afgA  
; OTHER INFORMATION: sequence containing the replacement fragment  
; OTHER INFORMATION: encoding PT3 from GP63 of Leishmania major.  
US-09-543-407-16

Query Match 73.8%; Score 577; DB 19; Length 151;  
Best Local Similarity 80.8%; Pred. No. 2.1e-53;  
Matches 122; Conservative 5; Mismatches 24; Indels 0; Gaps 0;  
QY 1 MKLLKVAFAAIVVSGSALAGVVPQWGGGNGHNGSSGPDSTLSIYQGSANAALQ 60  
Db 1 MKLLKVAFAAIVVSGSALAGVVPQWGGGNGHNGSSGPDSTLSIYQGSANAALQ 60  
QY 61 LVTRVVTHEMAHAGYNGADVGQADNSTIETQNGFRNNATIDQWNAKNSDITVGYGG 120  
Db 61 SDARKSETTITQSGYNGADVGQADNSTIETQNGFRNNATIDQWNAKNSDITVGYGG 120  
QY 121 NNAALVNQTASDSSVMVRQVGFNNATANQY 151  
Db 121 NNAALVNQTASDSSVMVRQVGFNNATANQY 151

## RESULT 14

US-09-543-407-7  
; Sequence 7, Application US/09543407  
; GENERAL INFORMATION:  
; APPLICANT: White, Aaron P.  
; APPLICANT: Doran, James L.  
; APPLICANT: Collinson, S. Karen  
; APPLICANT: Kay, William W.  
; TITLE OF INVENTION: BACTERIAL FIMBRIAL SYSTEM FOR  
; TITLE OF INVENTION: PRESENTATION OF HETEROLOGOUS PEPTIDE SEQUENCES  
; FILE REFERENCE: 920043.406  
; CURRENT APPLICATION NUMBER: US/09/543,407  
; CURRENT FILING DATE: 2000-04-05  
; NUMBER OF SEQ ID NOS: 59  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 7  
; LENGTH: 151  
; TYPE: PRT  
; ORGANISM: Escherichia coli  
US-09-543-407-7

Query Match 66.9%; Score 523; DB 19; Length 151;  
Best Local Similarity 68.9%; Pred. No. 1.4e-47;  
Matches 104; Conservative 18; Mismatches 29; Indels 0; Gaps 0;  
QY 1 MKLLKVAFAAIVVSGSALAGVVPQWGGGNGHNGSSGPDSTLSIYQGSANAALQ 60  
Db 1 MKLLKVAFAAIVVSGSALAGVVPQWGGGNGHNGSSGPDSTLSIYQGSANAALQ 60  
QY 61 LVTRVVTHEMAHAGYNGADVGQADNSTIETQNGFRNNATIDQWNAKNSDITVGYGG 120  
Db 61 TDARNSDLTITQHGNGGADVGQSDSSIDLTORFGNSATLDQWNGKNSMTVKQFGG 120  
QY 121 NNAALVNQTASDSSVMVRQVGFNNATANQY 151  
Db 121 GNGRAVDQTASNSSVNTQVGFNNATANQY 151

## RESULT 15

US-08-978-878-4  
; Sequence 4, Application US/08978878  
; GENERAL INFORMATION:  
; APPLICANT: NORVARK, Staffan  
; APPLICANT: OLSEN, Arne  
; TITLE OF INVENTION: FIBRONECTIN BINDING PROTEIN AS WELL AS ITS PREPARATION  
; FILE REFERENCE: 012889-081  
; CURRENT APPLICATION NUMBER: US/08/978,878  
; CURRENT FILING DATE: 1997-11-26  
; EARLIER APPLICATION NUMBER: SE 8801723-1  
; EARLIER FILING DATE: 1988-05-06

Wed Aug 4 10:14:31 2004

```
; EARLIER APPLICATION NUMBER: US 07/347,189
; EARLIER FILING DATE: 1989-05-04
; EARLIER APPLICATION NUMBER: US 07/789,437
; EARLIER FILING DATE: 1991-11-06
; EARLIER APPLICATION NUMBER: US 07/970,846
; EARLIER FILING DATE: 1992-11-03
; EARLIER APPLICATION NUMBER: US 08/187,865
; EARLIER FILING DATE: 1994-01-28
; EARLIER APPLICATION NUMBER: US 08/318,519
; EARLIER FILING DATE: 1994-10-05
; EARLIER APPLICATION NUMBER: US 08/495,959
; EARLIER FILING DATE: 1995-06-28
; NUMBER OF SEQ ID NOS: 10
; SOFTWARE: Patentin Ver. 2.0
; SEQ ID NO 4
; LENGTH: 151
; TYPE: PRT
; ORGANISM: Escherichia coli
US-08-978-878-4

Query Match      66.5%; Score 520; DB 13; Length 151;
Best Local Similarity 68.2%; Pred. No. 2.9e-47;
Matches 103; Conservative 19; Mismatches 29; Indels 0; Gaps 0;

QY      1 MLLKVAFAAIVVSGSALAGVVPQGGGNGHNGGNSGDPDSTLSIYQYGSANAALYDQ 60
Db      1 MLLKVAFAAIVFSGSAVAGVVPQGGGNGHNGGNSGDPDSTLSIYQYGSANAALYDQ 60

QY      61 LVTRVVTHEMAHAGYNGADVGQGDNSTIETQNGFRNATIDOWNAKNSDITVQYGG 120
Db      61 TDARNSDLITQHGCGGADVGQGDSDSIDLTQRFNGSATLDQNGKNSMTVKQFEG 120

QY      121 NNAALVNCTASDSSVMVROVFGNNATANOY 151
Db      121 GNGAAVDQTASNSVNVTVQVFGNNATAHOY 151
```

Search completed: August 2, 2004, 15:26:46  
Job time : 188.9 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: August 2, 2004, 14:49:38 ; Search time 17.8 Seconds

(without alignments)  
888.146 Million cell updates/sec

Title: US-09-543-407-26

Perfect score: 782

Sequence: 1 MKLLKVAFAAIVVSGSALA.....DSSVMVRQVGFNNATANQY 151

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 601315 seqs, 104695340 residues

Total number of hits satisfying chosen parameters: 601315

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%  
Listing first 45 summaries

Database : Pending Patents AA New.\*

1: /cgn2\_6/prodata/2/paa/PCT\_NEW\_COMB.pcp.\*  
2: /cgn2\_6/prodata/2/paa/US06\_NEW\_COMB.pcp.\*  
3: /cgn2\_6/prodata/2/paa/US07\_NEW\_COMB.pcp.\*  
4: /cgn2\_6/prodata/2/paa/US08\_NEW\_COMB.pcp.\*  
5: /cgn2\_6/prodata/2/paa/US09\_NEW\_COMB.pcp.\*  
6: /cgn2\_6/prodata/2/paa/US10\_NEW\_COMB.pcp.\*  
7: /cgn2\_6/prodata/2/paa/US60\_NEW\_COMB.pcp.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	520	66.5	151	5	US-09-741-873C-4
2	442	56.5	131	5	US-09-741-873C-2
3	93	11.9	892	5	US-09-952-267B-5
4	93	11.9	892	6	US-10-872-768-5
5	93	11.9	892	6	US-10-872-769-5
6	92	11.8	295	6	US-10-425-115-312468
7	92	11.8	1459	7	US-10-482-706-129
8	91.5	11.7	891	7	US-60-566-425-574
9	91.5	11.7	891	7	US-60-576-812-606
10	91.5	11.7	894	7	US-60-566-425-571
11	91.5	11.7	894	7	US-60-566-425-570
12	91.5	11.7	894	7	US-60-566-425-575
13	91.5	11.7	894	7	US-60-566-425-578
14	91.5	11.7	894	7	US-60-576-812-602
15	91.5	11.7	894	7	US-60-576-812-603
16	91.5	11.7	894	7	US-60-576-812-607
17	91.5	11.7	894	7	US-60-576-812-610
18	91.5	11.7	898	6	US-10-170-205B-12329
19	91.5	11.7	898	7	US-60-566-425-577
20	91.5	11.7	898	7	US-60-576-812-609
21	89	11.4	197	6	US-10-425-115-304391
22	88.5	11.3	193	6	US-10-425-115-254240
23	88	11.3	1627	6	US-10-170-205E-16659
24	88	11.3	1905	1	PCT-US04-09388-9
25	87.5	11.2	234	6	US-10-767-701-45603
26	87	11.1	573	7	US-60-565-632-7907

27 11.1 573 7 US-60-579-062-7907 Sequence 7907, Ap  
28 85.5 10.9 258 6 US-10-425-115-300390 Sequence 300390,  
29 85 10.9 147 6 US-10-425-115-193207 Sequence 193207,  
30 30 85 10.9 412 7 US-60-565-632-7905 Sequence 7905, Ap  
31 85 10.9 412 7 US-60-579-062-7905 Sequence 7905, Ap  
32 85 10.9 841 7 US-60-565-632-7906 Sequence 7906, Ap  
33 85 10.9 841 7 US-60-579-062-7906 Sequence 7906, Ap  
34 83.5 10.7 443 6 US-10-100-683-7608 Sequence 7608, Ap  
35 82 10.5 131 7 US-60-565-632-11109 Sequence 11109, A  
36 82 10.5 131 7 US-60-579-062-11109 Sequence 11109, A  
37 82 10.5 193 6 US-10-425-115-219256 Sequence 219256,  
38 82 10.5 299 6 US-10-170-205B-35751 Sequence 35751, A  
39 82 10.5 434 7 US-60-565-632-9168 Sequence 9168, Ap  
40 82 10.5 434 7 US-60-579-062-9168 Sequence 9168, Ap  
41 82 10.5 544 6 US-10-425-115-226699 Sequence 226699,  
42 82 10.5 1203 6 US-10-170-205E-741 Sequence 741, App  
43 82 10.5 1327 1 PCT-US04-02338-49 Sequence 49, Appl  
44 82 10.5 1358 6 US-10-778-804-11 Sequence 11, Appl  
45 81.5 10.4 179 6 US-10-767-701-35342 Sequence 35342, A

#### ALIGNMENTS

##### RESULT 1

US-09-741-873C-4  
; Sequence 4, Application US/09741873C  
; GENERAL INFORMATION:  
; APPLICANT: Normark, Staffan  
; TITLE OF INVENTION: Fibronectin Binding Protein As Well As Its Preparation  
; FILE REFERENCE: 012889-084  
; CURRENT APPLICATION NUMBER: US/09/741,873C  
; CURRENT FILING DATE: 2000-12-22  
; PRIOR APPLICATION NUMBER: SE 8801723-1  
; PRIOR FILING DATE: 1998-05-06  
; PRIOR APPLICATION NUMBER: US 08/978,878  
; PRIOR FILING DATE: 1997-11-26  
; PRIOR APPLICATION NUMBER: US 07/347,189  
; PRIOR FILING DATE: 1989-05-04  
; PRIOR APPLICATION NUMBER: US 07/789,437  
; PRIOR FILING DATE: 1991-11-06  
; PRIOR APPLICATION NUMBER: US 07/970,846  
; PRIOR FILING DATE: 1992-11-03  
; PRIOR APPLICATION NUMBER: US 08/187,865  
; PRIOR FILING DATE: 1994-01-28  
; PRIOR APPLICATION NUMBER: US 08/318,519  
; PRIOR FILING DATE: 1994-10-05  
; NUMBER OF SEQ ID NOS: 11  
; SOFTWARE: Patent in version 3.0  
; SEQ ID NO 4  
; LENGTH: 151  
; TYPE: PRT  
; ORGANISM: Escherichia coli  
US-09-741-873C-4

Query Match 66.5%; Score 520; DB 5; Length 151;  
Best Local Similarity 68.2%; Pred. No. 5.5e-38;  
Matches 103; Conservative 19; Mismatches 29; Indels 0; Gaps 0;

QY	1	MKLLKVAFAAIVVSGSALA	GVPGWGGGNGHNGSGSPDSTLSIYVYGSANAALYDQ	60
DB	1	MKLLKVAFAAIVVSGSALA	GVPGWGGGNGHNGSGSPDSTLSIYVYGSANAALYDQ	60
QY	61	LVTRYVTHEMAHAGYNGADYGGQADNSTIELTQNGFNATTIDQWNAKNSDITVQYGG	120	
DB	61	TDARNSDLITQHGCGNGADYGGQADNSTIELTQNGFNATTIDQWNAKNSDITVQYGG	120	
QY	121	NNAALVNOTASDSSVMVRQVGFNNATANQY	151	
DB	121	NGRAVDQTASNSSVMVRQVGFNNATANQY	151	

RESULT 2  
 US-09-741-873C-2  
 ; Sequence 2, Application US/09741873C  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Normark, Staffan  
 ; APPLICANT: Olsen, Arne  
 ; TITLE OF INVENTION: Fibronectin Binding Protein As Well As Its Preparation  
 ; FILE REFERENCE: 012889-084  
 ; CURRENT APPLICATION NUMBER: US/09/741,873C  
 ; CURRENT FILING DATE: 2000-12-22  
 ; PRIOR APPLICATION NUMBER: SE 8801723-1  
 ; PRIOR FILING DATE: 1998-05-06  
 ; PRIOR APPLICATION NUMBER: US 08/978,878  
 ; PRIOR FILING DATE: 1997-11-26  
 ; PRIOR APPLICATION NUMBER: US 07/347,189  
 ; PRIOR FILING DATE: 1989-05-04  
 ; PRIOR APPLICATION NUMBER: US 07/789,437  
 ; PRIOR FILING DATE: 1991-11-06  
 ; PRIOR APPLICATION NUMBER: US 07/970,846  
 ; PRIOR FILING DATE: 1992-11-03  
 ; PRIOR APPLICATION NUMBER: US 08/187,865  
 ; PRIOR FILING DATE: 1994-01-28  
 ; PRIOR APPLICATION NUMBER: US 08/318,519  
 ; PRIOR FILING DATE: 1994-10-05  
 ; NUMBER OF SEQ ID NOS: 11  
 ; SOFTWARE: Patent in version 3.0  
 ; SEQ ID NO 2  
 ; LENGTH: 131  
 ; TYPE: PRT  
 ; ORGANISM: Escherichia coli  
 US-09-741-873C-2

Query Match 56.5%; Score 442; DB 5; Length 131;  
 Best Local Similarity 64.9%; Pred. No. 2.8e-31;  
 Matches 85; Conservative 18; Mismatches 28; Indels 0; Gaps 0;  
 QY 21 GVVPQGGGNGHNGGSGPSTLSIYQYGSANAALYDQLVTRVVTTHMAHAGYGNAD 80  
 DB 1 GVVPQGGGNGHNGGSGPSELNTYQYGGNSALALQTDARNSDLTIHQHGGNGAD 60  
 QY 81 VGGADNSTLTQNGFRNATIDQNAKNSDITVGYGNNALYDQASDSSVMVQV 140  
 DB 61 VGGGSDSSIDLTRQGFSGSATLDWNGKNSMTVKQFGGNGAAVDQTPASNSVNTQV 120  
 QY 141 GFGNNATANQY 151  
 DB 121 GFGNNATANQY 131

RESULT 3  
 US-09-952-267B-5  
 ; Sequence 5, Application US/09952267B  
 ; GENERAL INFORMATION:  
 ; APPLICANT: HANSEN, ERIC J.  
 ; APPLICANT: AEBI, CHRISTOPH  
 ; APPLICANT: COPE, LESLIE D.  
 ; APPLICANT: MACIVER, ISOBEL  
 ; APPLICANT: FISKE, MICHAEL J.  
 ; APPLICANT: FREDENBURG, ROSS A.  
 ; TITLE OF INVENTION: USP1 AND USP2 ANTIGENS OF MORAXELLA CATARRHALIS  
 ; FILE REFERENCE: AMCY.024  
 ; CURRENT APPLICATION NUMBER: US/09/952,267B  
 ; CURRENT FILING DATE: 2001-09-12  
 ; PRIOR APPLICATION NUMBER: US/09/336,447  
 ; PRIOR FILING DATE: 1999-06-21  
 ; NUMBER OF SEQ ID NOS: 98  
 ; SOFTWARE: Patent in Ver. 2.1  
 ; SEQ ID NO 5  
 ; LENGTH: 892  
 ; TYPE: PRT  
 ; ORGANISM: Moraxella catarrhalis  
 US-09-952-267B-5

Query Match 11.9%; Score 93; DB 5; Length 892;  
 Best Local Similarity 27.7%; Pred. No. 5.7;  
 Matches 41; Conservative 14; Mismatches 49; Indels 44; Gaps 9;  
 QY 28 GGGNHN-----GGGNS---GPDSTLSIYQYGSANAALYDQLVTRVVTTHMAHAGY 75  
 DB 89 GGGDYNEAKGNYSTVGGSSNTAKGKSTIGGDTNDANG-----TVSTIGGGY 137  
 QY 76 -----GNGADVGGADNSTI--ELTQNGFRNATIDQNAKNSDITVGYG---GNNAL 125  
 DB 138 YSRAIGDSSTIGGYNQATGCKSTVAGGRN---QATGNNSTVAGGSGYNQATGNNSTV 193  
 QY 126 V---NQTASDSSVMVQVGFNNATAN 149  
 DB 194 AGGSHNQATGEGSF---AAGVENKANAN 218

RESULT 4  
 US-10-872-768-5  
 ; Sequence 5, Application US/10872768  
 ; GENERAL INFORMATION:  
 ; APPLICANT: HANSEN, ERIC J.  
 ; APPLICANT: AEBI, CHRISTOPH  
 ; APPLICANT: COPE, LESLIE D.  
 ; APPLICANT: MACIVER, ISOBEL  
 ; APPLICANT: FISKE, MICHAEL J.  
 ; APPLICANT: FREDENBURG, ROSS A.  
 ; TITLE OF INVENTION: USP1 AND USP2 ANTIGENS OF MORAXELLA CATARRHALIS  
 ; FILE REFERENCE: AMCY.024  
 ; CURRENT APPLICATION NUMBER: US/10/872,768  
 ; CURRENT FILING DATE: 2004-06-21  
 ; PRIOR APPLICATION NUMBER: US/09/336,447  
 ; PRIOR FILING DATE: 1999-06-21  
 ; NUMBER OF SEQ ID NOS: 98  
 ; SOFTWARE: Patent in Ver. 2.1  
 ; SEQ ID NO 5  
 ; LENGTH: 892  
 ; TYPE: PRT  
 ; ORGANISM: Moraxella catarrhalis  
 US-10-872-768-5

Query Match 11.9%; Score 93; DB 6; Length 892;  
 Best Local Similarity 27.7%; Pred. No. 5.7;  
 Matches 41; Conservative 14; Mismatches 49; Indels 44; Gaps 9;  
 QY 28 GGGNHN-----GGGNS---GPDSTLSIYQYGSANAALYDQLVTRVVTTHMAHAGY 75  
 DB 89 GGGDYNEAKGNYSTVGGSSNTAKGKSTIGGDTNDANG-----TVSTIGGGY 137  
 QY 76 -----GNGADVGGADNSTI--ELTQNGFRNATIDQNAKNSDITVGYG---GNNAL 125  
 DB 138 YSRAIGDSSTIGGYNQATGCKSTVAGGRN---QATGNNSTVAGGSGYNQATGNNSTV 193  
 QY 126 V---NQTASDSSVMVQVGFNNATAN 149  
 DB 194 AGGSHNQATGEGSF---AAGVENKANAN 218

RESULT 5  
 US-10-872-769-5  
 ; Sequence 5, Application US/10872769  
 ; GENERAL INFORMATION:  
 ; APPLICANT: HANSEN, ERIC J.  
 ; APPLICANT: AEBI, CHRISTOPH  
 ; APPLICANT: COPE, LESLIE D.  
 ; APPLICANT: MACIVER, ISOBEL  
 ; APPLICANT: FISKE, MICHAEL J.  
 ; APPLICANT: FREDENBURG, ROSS A.  
 ; TITLE OF INVENTION: USP1 AND USP2 ANTIGENS OF MORAXELLA CATARRHALIS  
 ; FILE REFERENCE: AMCY.024  
 ; CURRENT APPLICATION NUMBER: US/10/872,769  
 ; CURRENT FILING DATE: 2004-06-21  
 ; PRIOR APPLICATION NUMBER: US/09/336,447





```
; TITLE OF INVENTION: COLON DISEASE TARGETS AND USES THEREOF
; FILE REFERENCE: CL001527PROV
; CURRENT APPLICATION NUMBER: US/60/576,812
; CURRENT FILING DATE: 2004-06-04
; NUMBER OF SEQ ID NOS: 1501
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 606
; LENGTH: 891
; TYPE: PR1
; ORGANISM: Homo sapiens
US-60-576-812-606
```

```
Query Match 11.7%; Score 91.5; DB 7; Length 891;
Best Local Similarity 27.1%; Pred. No. 7.7;
Matches 35; Conservative 15; Mismatches 32; Indels 47; Gaps 7;

Qy 27 GGGG-----NHHGGGSSGPDSTLSIYQYGSANAALYDQVTRVVTHEMAHAGYGN 77
Db 792 GGGSDYNYESKFNYSGGSGSGNS-----YGGGASY-----NPGSHGGYGG 835

Qy 78 GADVGGQADNSTIELTQNGFRNNATIDQWNAKNSDITVQYGGNNAALVNQTASDSSVMV 137
Db 836 GSGGG-----SSYQKGQGGYSQS-----NYSNPGS-----CQ-----NYSGPPSSYSQS 873

Qy 138 ROVGFQGNNA 146
Db 874 SQGGYGRNA 882
```

```
RESULT 10
US-60-566-425-570
; Sequence 570, Application US/60566425
; GENERAL INFORMATION:
; APPLICANT: DOMON, Bruno et al.
; TITLE OF INVENTION: COLON DISEASE TARGETS AND USES THEREOF
; FILE REFERENCE: CL001522-PROV
; CURRENT APPLICATION NUMBER: US/60/566,425
; CURRENT FILING DATE: 2004-04-30
; NUMBER OF SEQ ID NOS: 1422
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 570
; LENGTH: 894
; TYPE: PR1
; ORGANISM: Homo sapiens
US-60-566-425-570
```

```
Query Match 11.7%; Score 91.5; DB 7; Length 894;
Best Local Similarity 27.1%; Pred. No. 7.7;
Matches 35; Conservative 15; Mismatches 32; Indels 47; Gaps 7;

Qy 27 GGGG-----NHHGGGSSGPDSTLSIYQYGSANAALYDQVTRVVTHEMAHAGYGN 77
Db 795 GGGSDYNYESKFNYSGGSGSGNS-----YGGGASY-----NPGSHGGYGG 838

Qy 78 GADVGGQADNSTIELTQNGFRNNATIDQWNAKNSDITVQYGGNNAALVNQTASDSSVMV 137
Db 839 GSGGG-----SSYQKGQGGYSQS-----NYSNPGS-----CQ-----NYSGPPSSYSQS 876

Qy 138 ROVGFQGNNA 146
Db 877 SQGGYGRNA 885
```

```
RESULT 11
US-60-566-425-571
; Sequence 571, Application US/60566425
; GENERAL INFORMATION:
; APPLICANT: DOMON, Bruno et al.
; TITLE OF INVENTION: COLON DISEASE TARGETS AND USES THEREOF
; FILE REFERENCE: CL001522-PROV
; CURRENT APPLICATION NUMBER: US/60/566,425
; CURRENT FILING DATE: 2004-04-30
; NUMBER OF SEQ ID NOS: 1422
```

```
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 571
; LENGTH: 894
; TYPE: PR1
; ORGANISM: Homo sapiens
US-60-566-425-571
```

```
Query Match 11.7%; Score 91.5; DB 7; Length 894;
Best Local Similarity 27.1%; Pred. No. 7.7;
Matches 35; Conservative 15; Mismatches 32; Indels 47; Gaps 7;

Qy 27 GGGG-----NHHGGGSSGPDSTLSIYQYGSANAALYDQVTRVVTHEMAHAGYGN 77
Db 795 GGGSDYNYESKFNYSGGSGSGNS-----YGGGASY-----NPGSHGGYGG 838

Qy 78 GADVGGQADNSTIELTQNGFRNNATIDQWNAKNSDITVQYGGNNAALVNQTASDSSVMV 137
Db 839 GSGGG-----SSYQKGQGGYSQS-----NYSNPGS-----CQ-----NYSGPPSSYSQS 876

Qy 138 ROVGFQGNNA 146
Db 877 SQGGYGRNA 885
```

```
RESULT 12
US-60-566-425-575
; Sequence 575, Application US/60566425
; GENERAL INFORMATION:
; APPLICANT: DOMON, Bruno et al.
; TITLE OF INVENTION: COLON DISEASE TARGETS AND USES THEREOF
; FILE REFERENCE: CL001522-PROV
; CURRENT APPLICATION NUMBER: US/60/566,425
; CURRENT FILING DATE: 2004-04-30
; NUMBER OF SEQ ID NOS: 1422
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 575
; LENGTH: 894
; TYPE: PR1
; ORGANISM: Homo sapiens
US-60-566-425-575
```

```
Query Match 11.7%; Score 91.5; DB 7; Length 894;
Best Local Similarity 27.1%; Pred. No. 7.7;
Matches 35; Conservative 15; Mismatches 32; Indels 47; Gaps 7;

Qy 27 GGGG-----NHHGGGSSGPDSTLSIYQYGSANAALYDQVTRVVTHEMAHAGYGN 77
Db 795 GGGSDYNYESKFNYSGGSGSGNS-----YGGGASY-----NPGSHGGYGG 838

Qy 78 GADVGGQADNSTIELTQNGFRNNATIDQWNAKNSDITVQYGGNNAALVNQTASDSSVMV 137
Db 839 GSGGG-----SSYQKGQGGYSQS-----NYSNPGS-----CQ-----NYSGPPSSYSQS 876

Qy 138 ROVGFQGNNA 146
Db 877 SQGGYGRNA 885
```

```
RESULT 13
US-60-566-425-578
; Sequence 578, Application US/60566425
; GENERAL INFORMATION:
; APPLICANT: DOMON, Bruno et al.
; TITLE OF INVENTION: COLON DISEASE TARGETS AND USES THEREOF
; FILE REFERENCE: CL001522-PROV
; CURRENT APPLICATION NUMBER: US/60/566,425
; CURRENT FILING DATE: 2004-04-30
; NUMBER OF SEQ ID NOS: 1422
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 578
; LENGTH: 894
; TYPE: PR1
; ORGANISM: Homo sapiens
```



This Page Blank (uspto)

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: August 2, 2004, 14:39:53 ; Search time 9.4 Seconds  
(without alignments)  
1545.204 Million cell updates/sec

Title: US-09-543-407-26

Perfect score: 782

Sequence: 1 MKLLKVAFAAIVVSGSALA.....DSSVMVRQVGFNNATANQY 151

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 283366 seqs, 96191526 residues

Total number of hits satisfying chosen parameters: 283366

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

PIR 78:\*\*

1: PIR1:\*\*

2: PIR2:\*\*

3: PIR3:\*\*

4: PIR4:\*\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	692	88.5	151	2 JC6039	fimbrin protein ag
2	692	88.5	151	2 A10635	major curlin chain
3	593	66.9	151	2 S70788	curlin protein csg
4	501.5	64.1	152	2 D90806	curlin major subun
5	501.5	64.1	152	2 H85665	hypothetical prote
6	114.5	14.6	145	2 AD3143	conserved hypotet
7	114.5	14.6	145	2 H98144	hypothetical glyci
8	113	14.5	2174	2 E95965	curlin nucleator p
9	112	14.3	151	2 S70787	minor curlin subun
10	112	14.3	151	2 C90806	curlin minor chain
11	112	14.3	151	2 G85665	ice nucleation pro
12	107.5	13.7	1322	2 S07053	fimbrin protein ag
13	107	13.7	151	2 JC6040	nucleation compone
14	107	13.7	151	2 AH0635	ice nucleation act
15	106	13.6	1034	2 JC2143	probable PPE prote
16	105.5	13.5	645	2 F70825	probable PPE prote
17	105	13.4	590	2 E70946	probable PPE prote
18	104.5	13.4	1258	2 JQ0188	ice nucleation pro
19	102.5	13.1	552	2 D70604	probable PPE prote
20	102.5	13.1	1567	2 S11672	ice nucleation pro
21	102.5	13.1	1655	2 E97835	hypothetical prote
22	100	12.8	3300	2 D70575	probable PPE prote
23	97.5	12.5	1028	2 A56038	DNA-binding protei
24	97.5	12.5	1213	2 S16356	ovo protein - frui
25	97	12.4	615	2 E70663	probable PPE prote
26	96.5	12.3	447	2 G84687	outer membrane pro
27	95.5	12.2	1651	2 JC1340	cell surface antig
28	95	12.1	652	2 E97857	leishmanolysin (BC
29	94	12.0	599	2 B42049	

30 94 12.0 599 2 A44951  
31 93 11.9 434 2 E70768  
32 92.5 11.8 575 2 S35327  
33 92.5 11.8 639 2 C42049  
34 92.5 11.8 963 2 E70524  
35 92.5 11.8 1053 2 E70987  
36 92 11.8 354 2 E70663  
37 92 11.8 1436 2 E70520  
38 92 11.8 1748 2 S42136  
39 91.5 11.7 582 2 F70675  
40 91 11.6 678 2 A70762  
41 90.5 11.6 1200 1 SNPSO  
42 90.5 11.6 3716 2 E70969  
43 90 11.5 586 2 T26667  
44 90 11.5 1390 2 T14004  
45 89.5 11.4 1317 2 A54831

#### ALIGNMENTS

##### RESULT 1

JC6039  
fimbrin protein agfA precursor - Salmonella enteritidis  
C:Species: Salmonella enteritidis  
C:Date: 31-Dec-1996 #sequence revision 31-Dec-1996 #text\_change 08-Oct-1999  
C:Accession: JC6039; PQ6015; A44898  
R:Collinson, S.K.; Clouthier, S.C.; Doran, J.L.; Bansen, P.A.; Kay, W.W.  
J. Bacteriol. 178, 662-667, 1996  
A:Title: Salmonella enteritidis agfBAC operon encoding thin, aggregative fimbriae.  
A:Reference number: JC6039; MUID:96146512; PMID:8550497  
A:Accession: JC6039  
A:Molecule type: DNA  
A:Residues: 1-151 <COL>  
A:Cross-references: GB:U43280; NID:g1184712; PIDN:AAC43599.1; PID:g1184714  
A:Accession: PQ6015  
A:Molecule type: protein  
A:Residues: 21-52 <CO2>  
A:Experimental source: strain 27655-3b  
A:Note: the authors translated the codon ACG for residue 44 as Ile  
R:Collinson, S.K.; Emody, L.; Muller, K.H.; Trust, T.J.; Kay, W.W.  
J. Bacteriol. 173, 4773-4781, 1991  
A:Title: Purification and characterization of thin, aggregative fimbriae from Salmonella  
A:Reference number: A44898; MUID:91310586; PMID:1677357  
A:Contents: 27655  
A:Accession: A44898  
A>Status: preliminary  
A:Molecule type: protein  
A:Residues: 21-33 <CO3>  
A:Note: sequence extracted from NCBI backbone (NCBIP:45936)  
C:Genetics:  
A:Gene: agfA  
C:Function:  
A:Description: major component of thin aggregative fimbriae  
A:Note: fimbriae bind to fibronectin, plasminogen, tissue plasminogen activator  
C:Keywords: fimbrin  
F:1-20/Domain: signal sequence #status predicted <SIG>  
F:21-151/Product: fimbrin protein agfA #status experimental <MAT>

Query Match 88.5%; Score 692; DB 2; Length 151;

Best Local Similarity 90.7%; Pred. No. 3.6e-51;

Matches 137; Conservative 2; Mismatches 12; Indels 0; Gaps 0;

QY 1 MKLLKVAFAAIVVSGSALAGVVPQWGGGNGHNGSGSPDSTLSIYQGSANAALYDQ 60

Db 1 MKLLKVAFAAIVVSGSALAGVVPQWGGGNGHNGSGSPDSTLSIYQGSANAALALQ 60

QY 61 LVTVVTHMAHAGYGNAGDVGGADNSTLTQTGFRNNATIDQWNAKSDITVGYGG 120

Db 61 SDARKSETTITQSGYGNAGDVGGADNSTLTQTGFRNNATIDQWNAKSDITVGYGG 120

QY 121 NNAALVNTQATSDSSVMVRQVGFNNATANQY 151

Db 121 NNAALVNTASDSSVMVRQVGFNNATANQY 151

RESULT 2

A:Title: The Rpos sigma factor precursor [imported] - Salmonella enterica subsp. enterica serovar Typhimurium  
 A:Species: Salmonella enterica subsp. enterica serovar Typhi  
 A:Note: this species has also been called Salmonella typhi  
 C:Date: 09-Nov-2001 #sequence\_revision 09-Nov-2001 #text\_change 18-Nov-2002  
 C:Accession: A10635  
 R:Parkhill, J.; Dougan, G.; James, K.D.; Thomson, N.R.; Pickard, D.; Wain, J.; Churcher, C.; Parkhill, J.; Cronin, A.; Davis, P.; Davies, R.M.; Dowd, L.; White, N.; Farrar, S.; Moule, S.; O'Gaora, P.  
 Nature 413, 848-852, 2001  
 A:Authors: Parkhill, J.; Cronin, A.; Davis, P.; Davies, R.M.; Dowd, L.; White, N.; Farrar, S.; Moule, S.; O'Gaora, P.  
 A:Title: Complete genome sequence of a multiple drug resistant Salmonella enterica serovar Typhimurium  
 A:Reference number: AB0502; MUID:21534947; PMID:11677608  
 A:Accession: A10635  
 A:Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-151 <PAR>  
 A:Cross-references: GB:AL513382; PIDN:CAD08268.1; PID:gl65023115; GSPDB:GN00176  
 C:Genetics:  
 A:Gene: STY1181

Query Match 88.5%; Score 692; DB 2; Length 151;  
 Best Local Similarity 90.7%; Pred. No. 3.6e-51;  
 Matches 137; Conservative 2; Mismatches 12; Indels 0; Gaps 0;

QY 1 MKLLKVAFAAIVVSGSALAGVVPQGGGNGGNGSGPSTLSIYVGSANAALYDQ 60  
 DB 1 MKLLKVAFAAIVVSGSALAGVVPQGGGNGGNGSGPSTLSIYVGSANAALQ 60

QY 61 LVTRVVTHEMAHAGYNGADVGQGDNSTIETQNGFRNNATIDQWNAKNSDITVQYGG 120  
 DB 61 SDARKSETTITQGGYNGADVGQGDNSTIETQNGFRNNATIDQWNAKNSDITVQYGG 120

QY 121 NNAALVNTASDSSVMVRQVGFNNATANQY 151  
 DB 121 NNAALVNTASDSSVMVRQVGFNNATANQY 151

RESULT 3

curlin protein csgA precursor - Escherichia coli (strain K-12)

N:Alternate names: csgA protein; major curlin protein  
 C:Species: Escherichia coli  
 C:Date: 12-Feb-1998 #sequence\_revision 20-Feb-1998 #text\_change 01-Mar-2002  
 C:Accession: S70788; G64846; S31202; S34560; S34559  
 R:Hammar, M.; Arngqvist, A.; Bian, Z.; Olsen, A.; Normark, S.  
 Mol. Microbiol. 18, 661-670, 1995  
 A:Title: Expression of two csg operons is required for production of fibronectin- and curli  
 A:Reference number: S70783; MUID:96414468; PMID:8817489  
 A:Accession: S70788  
 A:Status: nucleic acid sequence not shown; translation not shown  
 A:Molecule type: DNA  
 A:Residues: 1-151 <HAM>  
 A:Cross-references: EMBL:X00754; NID:gl147558; PIDN:CAA62282.1; PID:gl147564  
 A:Experimental source: strain K12, substrain W3110  
 A:Note: the nucleotide sequence was submitted to the EMBL Data Library, August 1995  
 R:Blattner, F.R.; Plunkett III, G.; Bloch, C.A.; Perna, N.T.; Burland, V.; Riley, M.; Cohen, D.; Rose, D.J.; Mau, B.; Shao, Y.  
 Science 277, 1453-1462, 1997  
 A:Title: The complete genome sequence of Escherichia coli K-12.  
 A:Reference number: A64720; MUID:57426617; PMID:9278503  
 A:Accession: G64846  
 A:Status: nucleic acid sequence not shown; translation not shown  
 A:Molecule type: DNA  
 A:Residues: 1-151 <BLAT>  
 A:Cross-references: GB:AE000205; GB:U00096; NID:gl1787265; PIDN:AACT4126.1; PID:gl1787279;  
 A:Experimental source: strain K-12, substrain MG1655  
 R:Olsen, A.; Arngqvist, A.; Hammar, M.; Sukupolvi, S.; Normark, S.  
 Mol. Microbiol. 7, 523-536, 1993

A:Title: The Rpos sigma factor relieves H-NS-mediated transcriptional repression of csgA,  
 A:Reference number: S31202; MUID:93211294; PMID:8459772  
 A:Accession: S31202  
 A:Molecule type: DNA  
 A:Residues: 1-6, 'V', 8-151 <OLS1>  
 A:Cross-references: EMBL:L04979  
 A:Accession: S34560  
 A:Molecule type: protein  
 A:Residues: 21-42; 44-50 <OLS2>  
 R:Olsen, A.N.; Arngqvist, A.M.  
 submitted to the EMBL Data Library, October 1992  
 A:Reference number: S34559  
 A:Accession: S34559  
 A:Molecule type: DNA  
 A:Residues: 1-133, 'RQRSGWLV' <OLS3>  
 A:Cross-references: EMBL:L04979; NID:g290424; PIDN:AAA23616.1; PID:g290425  
 A:Experimental source: strain K-12, substrain W3110  
 C:Genetics:  
 A:Gene: csgA  
 A:Map position: 23.15  
 C:Function:  
 A:Description: major component of wild-type curli; interaction between CsgA and CsgB tri-  
 A:Note: curli are thin, coiled fibers expressed on the surface of Escherichia coli that i  
 and H-kininogen; in the absence of CsgA, CsgB can self-assemble into polymers  
 F:1-20/Domain: signal sequence #status predicted <SIG>  
 F:21-151/Product: curlin #status experimental <MAT>

Query Match 66.9%; Score 523; DB 2; Length 151;  
 Best Local Similarity 68.9%; Pred. No. 5.3e-37;  
 Matches 104; Conservative 18; Mismatches 29; Indels 0; Gaps 0;

QY 1 MKLLKVAFAAIVVSGSALAGVVPQGGGNGGNGSGPSTLSIYVGSANAALYDQ 60  
 DB 1 MKLLKVAFAAIVVSGSALAGVVPQGGGNGGNGSGPSTLSIYVGSANAALQ 60

QY 61 LVTRVVTHEMAHAGYNGADVGQGDNSTIETQNGFRNNATIDQWNAKNSDITVQYGG 120  
 DB 61 TDARNSDLTITQGGGNGADVGQGDSSIDLQRFNGSATLDQWNGKNSMTVKQFG 120

QY 121 NNAALVNTASDSSVMVRQVGFNNATANQY 151  
 DB 121 GNGAADVDTASNSVNVTVQFGNNATAHQ 151

RESULT 4

curlin major subunit CsgA [imported] - Escherichia coli (strain O157:H7, substrain RIMD (

C:Species: Escherichia coli  
 C:Date: 18-Jul-2001 #sequence\_revision 18-Jul-2001 #text\_change 18-Jul-2001  
 C:Accession: D90806  
 R:Havashi, T.; Makino, K.; Kurokawa, K.; Ishii, K.; Yokoyama, K.; Han, C.G.;  
 Sasawara, N.; Yasunaga, T.; Kuhara, S.; Shiba, T.; Hattori, M.; Shinagawa, H.  
 DNA Res. 8, 11-22, 2001  
 A:Title: Complete genome sequence of enterohemorrhagic Escherichia coli O157:H7 and genom  
 A:Reference number: A99629; MUID:21156231; PMID:11258796  
 A:Accession: D90806  
 A:Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-152 <HAY>  
 A:Cross-references: GB:BA000007; PIDN:BA34843.1; PID:gl33360880; GSPDB:GN00154  
 A:Experimental source: strain O157:H7, substrain RIMD 050952  
 C:Genetics:  
 A:Gene: ECS1420

Query Match 64.1%; Score 501.5; DB 2; Length 152;  
 Best Local Similarity 67.1%; Pred. No. 3.4e-35;  
 Matches 102; Conservative 19; Mismatches 30; Indels 1; Gaps 1;

QY 1 MKLLKVAFAAIVVSGSALAGVVPQGGGNGGNGSGPSTLSIYVGSANAALYD 59  
 DB 1 MKLLKVAFAAIVVSGSALAGVVPQGGGNGGNGSGPSTLSIYVGSANAALQ 60

QY 60 QLVTRVVTHEMAHAGYNGADVGQGDNSTIETQNGFRNNATIDQWNAKNSDITVQYGG 119

Db 61 QADARNSDLTITGHHGGAGADVQGGSDSSIDITQRFNGSATLDWNGKDSHMTVKQFG 120  
QY 120 GNNAAALVNTQASDVSSVMVQVGFNNATANQY 151  
Db 121 GNGAAVDQTASNSTVNVTVQVGFNNATAHQY 152  
RESULT 5  
H85665  
hypothetical protein csgA [imported] - Escherichia coli (strain O157:H7, substrain EDL933)  
C:Species: Escherichia coli  
C>Date: 16-Feb-2001 #sequence\_revision 16-Feb-2001 #text\_change 14-Sep-2001  
C:Accession: H85665  
R:Perna, N.T.; Plunkett III, G.; Burland, V.; Mau, B.; Glasner, J.D.; Rose, D.J.; Mayhew, M.W.; Miller, L.; Grotbeck, E.J.; Davis, N.W.; Lim, A.; Dimalanta, E.; Potamousis, K.; Apodaca, Nature 409, 529-533, 2001  
A>Title: Genome sequence of enterohemorrhagic Escherichia coli O157:H7.  
A:Reference number: A85480; MUID:21074935; PMID:11206551  
A:Accession: H85665  
A>Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-152 <STO>  
A:Cross-references: GB:AE005174; NID:gl2514574; PIDN:AA055788.1; GSPDB:GN00145; UMGF:Z16  
A:Experimental source: strain O157:H7, substrain EDL933  
C:Genetics:  
A:Gene: csgA  
Query Match 64.1%; Score 501.5; DB 2; Length 152;  
Best Local Similarity 67.1%; Pred. No. 3.4e-35;  
Matches 102; Conservative 19; Mismatches 30; Indels 1; Gaps 1;  
QY 1 MKLLKVAFAAIVVSGSALAGVVPQW-GGGGHHNGGSSGPDSTLSIYQYGSANAALYD 59  
Db 1 MKLLKVAATAIVFSGSALAGVVPQYGGGGHHNGGSSGPNSEINIIYQYGGNSALAL 60  
QY 60 QLVTRVVTTHMAHAGYNGADVQGGADNSTIETQNGFRNNATIDQWNAKNSDITVQGVG 119  
Db 61 QADARNSDLTITGHHGGAGADVQGGSDSSIDITQRFNGSATLDWNGKDSHMTVKQFG 120  
QY 120 GNNAAALVNTQASDVSSVMVQVGFNNATANQY 151  
Db 121 GNGAAVDQTASNSTVNVTVQVGFNNATAHQY 152  
RESULT 6  
AD3143  
conserved hypothetical protein Atu4768 [imported] - Agrobacterium tumefaciens (strain C58)  
C:Species: Agrobacterium tumefaciens  
C>Date: 11-Jan-2002 #sequence\_revision 11-Jan-2002 #text\_change 18-Nov-2002  
C:Accession: AD3143  
R:Wood, D.W.; Setubal, J.C.; Kaul, R.; Monks, D.; Chen, L.; Wood, G.E.; Chen, Y.; Woo, I.; Karp, P.; Romero, P.; Zhang, S.  
Science 294, 2317-2323, 2001  
A:Authors: Yoo, H.; Tao, Y.; Biddle, P.; Jung, M.; Krespan, W.; Perry, M.; Gordon-Kamm, S.E.  
A>Title: The Genome of the Natural Genetic Engineer Agrobacterium tumefaciens C58.  
A:Reference number: AB2577; MUID:21608550; PMID:11743193  
A:Accession: AD3143  
A>Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-145 <KUR>  
A:Cross-references: GB:AE008699; PIDN:AA45562.1; PID:gl7743277; GSPDB:GN00187  
A:Experimental source: strain C58 (Dupont)  
C:Genetics:  
A:Gene: Atu4768  
A:Map position: linear chromosome  
Query Match 14.6%; Score 114.5; DB 2; Length 145;  
Best Local Similarity 25.2%; Pred. No. 0.0088;  
Matches 39; Conservative 26; Mismatches 59; Indels 31; Gaps 4;

QY 3 LLKVAFAAIVVSGSALAGVVPQW-----GGHHNGGSSGPDSTLSIYQYGSANA 55  
Db 1 MIRKSFASALVALVGLSAAAPAMANDVRIEYQYGSNSAGGAQYGNRIYQNGGYN- 59  
QY 56 ALYDQLVTRVVTTHMAHAGYNGADVQGGADNSTIETQNGFRNNATIDQWNAKNSDITV 115  
Db 60 -----RIVGHQY---GRHNL SAVQEGHDNVTGSTQNGRNVAGI----- 96  
QY 116 QYGGNNAALVNTQASDVSSVMVQVGFNNATANQ 150  
Db 97 QQFGSNHTTITQDNGNGNIAAGVQVGRGCSANVSQ 131  
RESULT 7  
H98144  
hypothetical protein AGR\_L\_228 [imported] - Agrobacterium tumefaciens (strain C58, Cereor)  
C:Species: Agrobacterium tumefaciens  
C>Date: 22-Oct-2001 #sequence\_revision 22-Oct-2001 #text\_change 18-Nov-2002  
C:Accession: H98144  
R:Goodner, B.; Hinkle, G.; Gattung, S.; Miller, N.; Blanchard, M.; Quorollo, B.; Goldman, A.; Liu, F.; Wollam, C.; Allinger, M.; Doughty, D.; Scott, C.; Lappas, C.; Markelz, B.; Science 294, 2323-2328, 2001  
A>Title: Genome Sequence of the Plant Pathogen and Biotechnology Agent Agrobacterium tumefaciens AGR\_L\_228.  
A:Reference number: A97359; MUID:21608551; PMID:11743194  
A:Accession: H98144  
A>Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-145 <KUR>  
A:Cross-references: GB:AE007870; PIDN:AAK88682.1; PID:gl5158413; GSPDB:GN00170  
C:Genetics:  
A:Gene: AGR\_L\_228  
A:Map position: linear chromosome  
Query Match 14.6%; Score 114.5; DB 2; Length 145;  
Best Local Similarity 25.2%; Pred. No. 0.0088;  
Matches 39; Conservative 26; Mismatches 59; Indels 31; Gaps 4;  
QY 3 LLKVAFAAIVVSGSALAGVVPQW-----GGHHNGGSSGPDSTLSIYQYGSANA 55  
Db 1 MIRKSFASALVALVGLSAAAPAMANDVRIEYQYGSNSAGGAQYGNRIYQNGGYN- 59  
QY 56 ALYDQLVTRVVTTHMAHAGYNGADVQGGADNSTIETQNGFRNNATIDQWNAKNSDITV 115  
Db 60 -----RIVGHQY---GRHNL SAVQEGHDNVTGSTQNGRNVAGI----- 96  
QY 116 QYGGNNAALVNTQASDVSSVMVQVGFNNATANQ 150  
Db 97 QQFGSNHTTITQDNGNGNIAAGVQVGRGCSANVSQ 131  
RESULT 8  
B95965  
hypothetical glycine-rich protein [imported] - Sinorhizobium meliloti (strain 1021) magat  
C:Species: Sinorhizobium meliloti  
C>Date: 24-Aug-2001 #sequence\_revision 24-Aug-2001 #text\_change 30-Sep-2001  
C:Accession: B95965  
R:Finan, T.M.; Weidner, S.; Wong, K.; Buhrmester, J.; Chain, P.; Vorholter, F.J.; Hernand Proc. Natl. Acad. Sci. U.S.A. 98, 9889-9894, 2001  
A>Title: The complete sequence of the 1,683-kb pSymB megaplasmid from the N2-fixing endo Science 294, 2317-2323, 2001  
A:Reference number: A95842; MUID:21396508; PMID:11481431  
A:Accession: B95965  
A>Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-2174 <KUR>  
A:Cross-references: GB:AL591985; PIDN:CAC49389.1; PID:gl5140875; GSPDB:GN00167  
A:Experimental source: strain 1021, megaplasmid pSymB  
R:Galibert, F.; Finan, T.M.; Long, S.R.; Puhler, A.; Abolia, P.; Ampe, F.; Barloy-Hubler, L.; Hyman, R.W.; Jones, T.; Science 293, 668-672, 2001  
A:Authors: Kahn, D.; Kahn, M.L.; Kalman, S.; Keating, D.H.; Kiss, E.; Komp, C.; Lelaure, hebault, P.; Vandenbol, M.; Vorholter, F.J.; Weidner, S.; Wells, D.H.; Wong, K.; Yeh, K. A:Title: The composite genome of the legume symbiont Sinorhizobium meliloti.

A:Reference number: A96039; MUID:21368234; PMID:11474104

A:Contents: annotation

C:Genetics:

A:Gene: SMD21548

A:Genome: plasmid

Query Match  
Best Local Similarity 14.5%; Score 113; DB 2; Length 2174;  
Matches 40; Conservative 21; Mismatches 53; Indels 36; Gaps 7;  
QY 11 AIVGSGALAGVVPQ--WGGGNGHNGGNSGPDSTLSIYQVGSANAALYDQVTRVTH 68  
DB 693 AATAGAGAVGILAQSIGGGGCGN--GGNATGDAGFGSGFQIGGGGGG----- 737  
QY 69 EWAHAGYNGADV-----QGADNSTI--ELTQNGFRNNATIDQNAK---NSDITV 115  
DB 738 ----GGYANTANVFKGLTLTTQGSAAAGIVAAQVGGGGTGGTASSVAGIGFTASVAV 793  
QY 116 GQVGGNAA--LVNQASDSSVMVRQV 141  
DB 794 GGTGGNGGAGGVSVLDSAIRTGQGG 821

RESULT 9  
S70787  
curlin nucleator protein csbB precursor - Escherichia coli (strain K-12)  
N:Alternate names: csbB protein; curlin nucleation component; minor curlin protein  
C:Species: Escherichia coli  
C:Date: 12-Feb-1998 #sequence\_revision 20-Feb-1998 #text\_change 01-Mar-2002  
C:Accession: S70787; F64846  
R:Hammar, M.; Arqvist, A.; Bian, Z.; Olsen, A.; Normark, S.  
Mol. Microbiol. 18, 661-670, 1995  
A:Title: Expression of two csb operons is required for production of fibronectin- and Cc  
A:Reference number: S70783; MUID:96414468; PMID:8817489  
A:Accession: S70787  
A:Status: nucleic acid sequence not shown; translation not shown  
A:Molecule type: DNA  
A:Residues: 1-151 <HAM>  
A:Cross-references: EMBL:X30754; NID:g1147558; PIDN:CAA62281.1; PID:g1147563  
A:Experimental source: strain K12, substrain W3110  
A:Note: the nucleotide sequence was submitted to the EMBL Data Library, August 1995  
R:Blattner, F.R.; Plunkett III, G.; Bloch, C.A.; Perna, N.T.; Burland, V.; Riley, M.; Cc  
A.; Rose, D.J.; Mau, B.; Shao, Y.  
Science 277, 1453-1462, 1997  
A:Title: The complete genome sequence of Escherichia coli K-12.  
A:Reference number: A64720; MUID:97426617; PMID:9278503  
A:Accession: F64846  
A:Status: nucleic acid sequence not shown; translation not shown  
A:Molecule type: DNA  
A:Residues: 1-151 <BLAT>  
A:Cross-references: GB:AE000205; NID:U00096; PIDN:AACT4125.1; PID:g1787278;  
A:Experimental source: strain K-12, substrain MGL655  
C:Genetics:  
A:Gene: csbB

A:Map position: 23.15  
C:Function:  
A:Description: minor component of wild-type curli; interaction between CsgA and CsgB tri  
A:Note: curli are thin, coiled fibers expressed on the surface of Escherichia coli that  
and H-kininogen; in the absence of CsgA, CsgB can self-assemble into polymers  
F:1-21/Domain: signal sequence #status predicted <SIG>  
F:22-151/Product: minor curlin chain #status predicted <MAT>

Query Match  
Best Local Similarity 14.3%; Score 112; DB 2; Length 151;  
Matches 34; Conservative 17; Mismatches 49; Indels 14; Gaps 3;  
QY 51 GSANAALYDQVLT--RVVTHEMAHAGYNGADVGGQADNSTIETLQNGFR----- 98  
DB 17 GIAAAGYDLANSEYFNFAVNELSKSFNQAAIIQAGTNNNAQLRQGGSKLLAVVAQEGS 76  
QY 99 -NNATIDQNAKSDITVGGYGGNNAALVNQTASDSSVMVRQVGFNNATANY 151  
DB 77 SNRAKIDQTDGYNL-AVIDQAGSANDASISQAGYNTAMIIQKSGNKANITQY 129

# RESULT 10

C90806  
minor curlin subunit precursor CsgB [imported] - Escherichia coli (strain O157:H7, subst  
C:Species: Escherichia coli  
C:Date: 18-Jul-2001 #sequence\_revision 18-Jul-2001 #text\_change 18-Jul-2001  
C:Accession: C90806  
R:Havashi, T.; Makino, K.; Ohnishi, M.; Kurokawa, K.; Ishii, K.; Yokoyama, K.; Han, C.G.;  
gasawara, N.; Yasunaga, T.; Kuhara, S.; Shiba, T.; Hattori, M.; Shinagawa, H.  
DNA Res. 8, 11-22, 2001  
A:Title: Complete genome sequence of enterohemorrhagic Escherichia coli O157:H7 and genom  
A:Reference number: A9629; MUID:21156231; PMID:11258796  
A:Accession: C90806  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-151 <HAY>  
A:Cross-references: GB:BA000007; PIDN:BA834842.1; PID:g13360879; GSPDB:GN00154  
A:Experimental source: strain O157:H7, substrain RIMD 0509952  
C:Genetics:  
A:Gene: ECS1419

Query Match 14.3%; Score 112; DB 2; Length 151;  
Best Local Similarity 29.8%; Pred. No. 0.015;  
Matches 34; Conservative 17; Mismatches 49; Indels 14; Gaps 3;

QY 51 GSANAALYDQVLT--RVVTHEMAHAGYNGADVGGQADNSTIETLQNGFR----- 98  
DB 17 GIAAAGYDLANSEYFNFAVNELSKSFNQAAIIQAGTNNNAQLRQGGSKLLAVVAQEGS 76  
QY 99 -NNATIDQNAKSDITVGGYGGNNAALVNQTASDSSVMVRQVGFNNATANY 151  
DB 77 SNRAKIDQTDGYNL-AVIDQAGSANDASISQAGYNTAMIIQKSGNKANITQY 129

# RESULT 11

G85665  
curlin minor chain precursor, CsgA homolog [imported] - Escherichia coli (strain O157:H7,  
C:Species: Escherichia coli  
C:Date: 16-Feb-2001 #sequence\_revision 16-Feb-2001 #text\_change 14-Sep-2001  
C:Accession: G85665  
R:Perna, N.T.; Plunkett III, G.; Burland, V.; Mau, B.; Glasner, J.D.; Rose, D.J.; Mayhew,  
Miller, L.; Grobeck, E.J.; Davis, N.W.; Lim, A.; Dimalanta, E.; Potamoudis, K.; Apodaca,  
Nature 409, 529-533, 2001  
A:Title: Genome sequence of enterohemorrhagic Escherichia coli O157:H7.  
A:Reference number: A85480; MUID:21074935; PMID:11206551  
A:Accession: G85665  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-151 <STO>  
A:Cross-references: GB:AE005174; NID:g12514573; PIDN:AAG55787.1; GSPDB:GN00145; UWGP:Z16;  
A:Experimental source: strain O157:H7, substrain EDL933  
C:Genetics:  
A:Gene: CsgB

Query Match 14.3%; Score 112; DB 2; Length 151;  
Best Local Similarity 29.8%; Pred. No. 0.015;  
Matches 34; Conservative 17; Mismatches 49; Indels 14; Gaps 3;

QY 51 GSANAALYDQVLT--RVVTHEMAHAGYNGADVGGQADNSTIETLQNGFR----- 98  
DB 17 GIAAAGYDLANSEYFNFAVNELSKSFNQAAIIQAGTNNNAQLRQGGSKLLAVVAQEGS 76  
QY 99 -NNATIDQNAKSDITVGGYGGNNAALVNQTASDSSVMVRQVGFNNATANY 151  
DB 77 SNRAKIDQTDGYNL-AVIDQAGSANDASISQAGYNTAMIIQKSGNKANITQY 129

# RESULT 12

S07053  
ice nucleation protein inaA - Erwinia ananas  
C:Species: Erwinia ananas  
C:Date: 30-Sep-1991 #sequence\_revision 30-Sep-1991 #text\_change 26-Aug-1999



C:Accession: S07053  
R:Abe, K.; Watabe, S.; Emori, Y.; Watanabe, M.; Arai, S.  
FBS Lett. 258, 297-300, 1989  
A:Title: An ice nucleation active gene of *Erwinia ananas*. Sequence similarity to those of  
A:Reference number: S07053; MUID:90092494; PMID:2599095  
A:Accession: S07053  
A>Status: not compared with conceptual translation  
A:Molecule type: DNA  
A:Residues: 1-1322 <ABE>  
A:Cross-references: GB:X17316; NID:G296095; PIDN:CAA35194.1; PID:g296096  
C:Superfamily: ice nucleation protein

Query Match 13.7%; Score 107.5; DB 2; Length 1322;  
Best Local Similarity 29.9%; Pred. No. 0.38; Mismatches 19; Indels 49; Gaps 8;  
Matches 38; Conservative 19

QY 34 GCGNSGGPSTLSIYQYGSANAALYDQLVTRVVTTHMAHAGYNGADYVQCGADNSTIILT 93  
DB 933 GSTSTAGPSSLL-IAGYGSTQTAGYNSIILT-----AGYGS-----T 967

QY 94 QNGFRNATIDQWNAKNSDIT-VGQYGGNNAALVNTASDSSVMVQVFGNNTANQY 143  
DB 968 QTC-----QENSLLTTG-YGSTSTAGYESSLIAGYGSTQTASFKSTLM--AGYG 1013

QY 144 NNATANQ 150  
DB 1014 SSQTARE 1020

RESULT 13  
JC6040  
fimbria protein agfB precursor - *Salmonella enteritidis*  
C:Species: *Salmonella enteritidis*  
C:Date: 31-Dec-1996 #sequence\_revision 31-Dec-1996 #text\_change 08-Oct-1999  
C:Accession: JC6040  
R:Collinson, S.K.; Clouthier, S.C.; Doran, J.L.; Baner, P.A.; Kay, W.W.  
J. Bacteriol. 178, 662-667, 1996  
A:Title: *Salmonella enteritidis* agfBAC operon encoding thin, aggregative fimbriae.  
A:Reference number: JC6039; MUID:96146512; PMID:8550497  
A:Accession: JC6040  
A:Molecule type: DNA  
A:Residues: 1-151 <COL>  
A:Cross-references: GB:U43280; NID:g1184712; PIDN:AA043598.1; PID:g1184713  
A:Experimental source: strain 276755-3b  
C:Genetics:  
A:Gene: agfB  
C:Function:  
A:Description: minor component of thin aggregative fimbriae  
A:Note: fimbriae bind to fibronectin, plasminogen, tissue plasminogen activator  
C:Keywords: fimbria  
F:1-21/Domain: signal sequence #status predicted <SIG>  
F:22-151/Product: fimbria protein agfB #status predicted <MAT>

Query Match 13.7%; Score 107; DB 2; Length 151;  
Best Local Similarity 30.4%; Pred. No. 0.039;  
Matches 35; Conservative 17; Mismatches 47; Indels 16; Gaps 5;

QY 51 GSANAALYD--QLVTRVVTTHMAHAGYNGADYVQ--GADNST-----IELTQNGF 97  
DB 17 GIATATNYDLARSEYFAVNELSKSFNQAAIIGQVGTDSARVRQSGSKLLSVISQGG 76

QY 98 RNNATIDQWNAKNSDIT-VGQYGGNNAALVNTASDSSVMVQVFGNNTANQY 151  
DB 77 NNRKVDQ--AGNYNPAVIEQTGNANDASISQAYGNSAAIIQKSGCNKANITQY 129

RESULT 14  
AH0635  
nucleation component of curlin monomers [imported] - *Salmonella enterica* subsp. *enterica*  
C:Species: *Salmonella enterica* subsp. *enterica* serovar Typhi  
A:Note: this species has also been called *Salmonella typhi*  
C:Date: 09-Nov-2001 #sequence\_revision 09-Nov-2001 #text\_change 18-Nov-2002  
C:Accession: AH0635

R:Parkhill, J.; Dougan, G.; James, K.D.; Thomson, N.R.; Pickard, D.; Wain, J.; Churcher,  
th, T.; Connerton, P.; Cronin, A.; Davis, P.; Davies, R.M.; Dowd, L.; White, N.; Farrar,  
S.; Moule, S.; O'Gaora, P.  
Nature 413, 848-852, 2001  
A:Authors: Parry, C.; Quail, M.; Rutherford, K.; Simmonds, M.; Skelton, J.; Stevens, K.;  
A:Title: Complete genome sequence of a multiple drug resistant *Salmonella enterica* serovar  
A:Reference number: AB0502; MUID:21534947; PMID:11677608  
A:Accession: AH0635  
A>Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-151 <PAR>  
A:Cross-references: GB:AL513382; PIDN:CAD08267.1; PID:g16502314; GSPDB:GN00176  
C:Genetics:  
A:Gene: STY1180

Query Match 13.7%; Score 107; DB 2; Length 151;  
Best Local Similarity 30.4%; Pred. No. 0.039;  
Matches 35; Conservative 17; Mismatches 47; Indels 16; Gaps 5;

QY 51 GSANAALYD--QLVTRVVTTHMAHAGYNGADYVQ--GADNST-----IELTQNGF 97  
DB 17 GIATATNYDLARSEYFAVNELSKSFNQAAIIGQVGTDSARVRQSGSKLLSVISQGG 76

QY 98 RNNATIDQWNAKNSDIT-VGQYGGNNAALVNTASDSSVMVQVFGNNTANQY 151  
DB 77 NNRKVDQ--AGNYNPAVIEQTGNANDASISQAYGNSAAIIQKSGCNKANITQY 129

RESULT 15  
JC2143  
ice nucleation active protein - *Erwinia uredovora* (strain KUIN-3)  
C:Species: *Erwinia uredovora*  
C:Date: 30-Sep-1993 #sequence\_revision 20-Aug-1994 #text\_change 17-Mar-1999  
C:Accession: JC2143  
R:Michigami, Y.; Watabe, S.; Abe, K.; Obata, H.; Arai, S.  
Biosci. Biotechnol. Biochem. 58, 762-764, 1994  
A:Title: Cloning and sequencing of an ice nucleation active gene of *Erwinia uredovora*.  
A:Reference number: JC2143; MUID:94264407; PMID:7764866  
A:Accession: JC2143  
A:Molecule type: DNA  
A:Residues: 1-1034 <MIC>  
C:Comment: This protein consists of 52 repeats of closely related 16-amino acid motifs (F  
C:Superfamily: ice nucleation protein  
F:161-993/Region: R-domain

Query Match 13.6%; Score 106; DB 2; Length 1034;  
Best Local Similarity 29.9%; Pred. No. 0.39; Mismatches 19; Indels 38; Gaps 9;  
Matches 40; Conservative 19

QY 27 GCGNHNHGGNSGPDSTLSIYQYGSANAALYDQLVTRVVTTHMAHAGYNGADYVQCGAD 86  
DB 531 GYGSTTAGANSS-----LIAGYGSTQTASVNSVLT-----AGYGSTQTAREGSD 575

QY 87 NSTIELTQNGFRNATIDQWNAKNSDITVGOYG-----GNNAALV-----NOTASDSSVM 136  
DB 576 -----LTA-GYGSTQTAE-----NSDLTTG-YGSTSTAGYDSSLIAGYGSTQTAGYHSIL 624

QY 137 VRQVFGNNTANQ 150  
DB 625 T--AGYGSTQTAE 636

Search completed: August 2, 2004, 14:56:25  
Job time : 10.4 secs

This Page Blank (uspto)

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: August 2, 2004, 14:36:12 ; Search time 5.3 Seconds  
(without alignments)  
1483.508 Million cell updates/sec

Title: US-09-543-407-26  
Perfect score: 782  
Sequence: 1 MKLLKVAFAAIVVSSSALA.....DSSVMVROVGFNNATANOY 151

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 141681 seqs, 52070155 residues  
Total number of hits satisfying chosen parameters: 141681

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : SwissProt\_42.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	692	88.5	151	1	CSGA_SALTY
2	523	66.9	151	1	CSGB_ECOLI
3	501.5	64.1	152	1	CSGB_ECOLI
4	112	14.3	151	1	CSGB_ECOLI
5	107.5	13.7	1322	1	ICGB_PANAN
6	107	13.7	151	1	ICGB_SALTY
7	107	13.7	151	1	CSGB_SALTY
8	106	13.6	1034	1	ICEN_PANAN
9	104.5	13.4	1258	1	ICEN_EWHE
10	102.5	13.1	1567	1	ICEN_XANCT
11	102.5	13.1	1655	1	ICEN_RICCN
12	97.5	12.5	1028	1	OVO_DROME
13	96.5	12.3	1196	1	ICEV_PSESX
14	95.5	12.2	1656	1	OMPB_RICJA
15	94	12.0	599	1	GP63_LEICH
16	93	11.9	431	1	YK98_MYCTU
17	91.5	11.7	894	1	LFPS_HUMAN
18	91	11.6	678	1	YF48_MYCTU
19	90.5	11.6	1200	1	ICEN_PSEY
20	89.5	11.4	1317	1	N145_YEAST
21	88.5	11.3	646	1	GP63_LEIME
22	88.5	11.3	1148	1	ICEK_PSESX
23	88.5	11.3	1210	1	ICEN_PSEFL
24	87.5	11.2	602	1	GP63_LEIMA
25	86.5	11.1	590	1	GP63_LEIDO
26	86	11.0	959	1	N100_YEAST
27	85	10.9	681	1	YDHE_SCHPO
28	84.5	10.8	548	1	CEAK_ECOLI
29	84.5	10.8	641	1	IMD_ARTGO
30	84.5	10.8	1300	1	120K_RICRI
31	84.5	10.8	1654	1	OMPB_RICRI
32	82.5	10.5	487	1	Y442_MYCTU
33	82	10.5	1327	1	TNKL_HUMAN

ALIGNMENTS

RESULT 1	CSGA_SALTY	ID	CSGA_SALTY	STANDARD;	PRT;	151 AA.
AC	P55225;	DT	01-OCT-1996 (Rel. 34, Created)			
DT	01-OCT-1996 (Rel. 34, Last sequence update)					
DT	10-OCT-2003 (Rel. 42, Last annotation update)					
DE	Major curlin subunit precursor (Pimbrin SEF17).					
GN	CSGA OR AGFA OR STM1144 OR STY1181 OR T1776.					
OS	Salmonella typhimurium,					
OS	Salmonella typhi, and					
OS	Salmonella enteritidis.					
OC	Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;					
OC	Enterobacteriaceae; Salmonella.					
OX	NCBI_TaxID=602, 601, 592;					
RN	[1]					
RP	SEQUENCE FROM N.A.					
RC	SPECIES=S.typhimurium; STRAIN=SR-11;					
RX	MEDLINE=98117058; PubMed=9457880;					
RA	Romling U., Bian Z., Hammar M., Sierralta W.D., Normark S.;					
RT	"Curli fibers are highly conserved between Salmonella typhimurium and					
RT	Escherichia coli with respect to operon structure and regulation."					
RL	J. Bacteriol. 180:722-731(1998).					
RL	[2]					
RP	SEQUENCE FROM N.A.					
RC	SPECIES=S.typhimurium; STRAIN=LT2 / SGSC1412 / ATCC 700720;					
RX	MEDLINE=21534948; PubMed=11677609;					
RA	McClelland M., Sanderson K.E., Spieth J., Clifton S.W., Latreille P.,					
RA	Courtney L., Porwollik S., Ali J., Dente M., Du F., Hou S., Layman D.,					
RA	Leonard S., Nguyen C., Scott K., Holmes A., Grewal N., Mulvaney E.,					
RA	Ryan E., Sun H., Florea L., Miller W., Stoneking T., Nhan M.,					
RA	Waterston R., Wilson R.K.;					
RT	"Complete genome sequence of Salmonella enterica serovar Typhimurium					
RT	LT2."					
RL	Nature 413:852-856(2001).					
RL	[3]					
RP	SEQUENCE FROM N.A.					
RC	SPECIES=S.typhi; STRAIN=CT18;					
RX	MEDLINE=21534947; PubMed=11677608;					
RA	Parthill J., Dougan G., James K.D., Thomson N.R., Pickard D., Wain J.,					
RA	Churcher C., Mungall K.L., Bentley S.D., Holden M.T.G., Sebahia M.,					
RA	Baker S., Basham D., Brooks K., Chillingworth T., Conerton P.,					
RA	Cronin A., Davis P., Davies R.M., Dowd L., White N., Farrar J.,					
RA	Feltwell T., Hamlin N., Haque A., Hien T.T., Holroyd S., Jagers K.,					
RA	Krogh A., Larsen T.S., Leather S., Moule S., O'Gaora P., Parry C.,					
RA	Quail M.A., Rutherford K., Simmonds M., Skelton J., Stevens K.,					
RA	Whitehead S., Barrall B.G.;					
RT	"Complete genome sequence of a multiple drug resistant Salmonella					
RT	enterica serovar Typhi CT18."					
RL	Nature 413:848-852(2001).					
RL	[4]					
RP	SEQUENCE FROM N.A.					
RC	SPECIES=S.typhi; STRAIN=TY2 / ATCC 700931;					
RX	MEDLINE=22531367; PubMed=12644504;					
RA	Deng W., Liou S.-R., Plunkett G. III, Mayhew G.F., Rose D.J.,					
RA	Burland V., Kodoyianni V., Schwartz D.C., Blattner P.R.;					

Q03646 plasmodium  
Q9utk4 schizosacch  
P07909 drosophila  
Q10540 mycobacteri  
Q8w234 arabidopsis  
P27692 saccharomyc  
P24785 drosophila  
P18431 drosophila  
P25074 oryza sativ  
P51989 xenopus lae  
Q9k039 neisseria m  
P96989 r outer mem

"Comparative genomics of *Salmonella enterica* serovar Typhi strains Ty2 and CT18.";  
 J. Bacteriol. 185:2330-2337(2003).  
 [5]  
 SEQUENCE FROM N.A.  
 SPECIES=S.enteritidis; STRAIN=27655-3B;  
 MEDLINE=96146512; PubMed=8550497;  
 Collinson S.K., Clouthier S.C., Doran J.L., Baner P.A., Kay W.W.;  
 "Salmonella enteritidis agrBAC operon encoding thin, aggregative  
 fimbriae.";  
 J. Bacteriol. 178:662-667(1996).  
 [6]  
 SEQUENCE OF 21-151 FROM N.A.  
 SPECIES=S.enteritidis; STRAIN=27655-3B;  
 MEDLINE=94013373; PubMed=8104955;  
 Doran J.L., Collinson S.K., Burian J., Santos G., Todd E.C.D.,  
 Munro C.K., Kay W.W., Baner P.A., Peterkin P.I., Kay W.W.;  
 "DNA-based diagnostic tests for *Salmonella* species targeting *agfA*,  
 the structural gene for thin, aggregative fimbriae.";  
 J. Clin. Microbiol. 31:2263-2273(1993).  
 [7]  
 SEQUENCE OF 21-33.  
 SPECIES=S.enteritidis; STRAIN=27655-3B;  
 MEDLINE=91310586; PubMed=1677357;  
 Collinson S.K., Emeody L., Mueller K.-M., Trust T.J., Kay W.W.;  
 "Purification and characterization of thin, aggregative fimbriae from  
*Salmonella enteritidis*.";  
 J. Bacteriol. 173:4773-4781(1991).  
 CC -1- FUNCTION: CURLIN IS THE STRUCTURAL SUBUNIT OF THE CURLI. CURLI ARE  
 COILED SURFACE STRUCTURES THAT ASSEMBLE PREFERENTIALLY AT GROWTH  
 TEMPERATURES BELOW 37 DEGREES CELSIUS. CURLI CAN BIND TO  
 FIBRONECTIN.  
 CC -1- SIMILARITY: BELONGS TO THE CSGA/CSGB FAMILY.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 between the Swiss Institute of Bioinformatics and the EMBL outstation  
 at the European Bioinformatics Institute. There are no restrictions on its  
 use by non-profit institutions as long as its content is in no way  
 modified and this statement is not removed. Usage by and for commercial  
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL; AJ002301; CAA05317.1; -;  
 DR EMBL; AB008749; AAL20074.1; -;  
 DR EMBL; AL627269; CAD08268.1; -;  
 DR EMBL; AE016840; AAC069399.1; -;  
 DR EMBL; U43280; AAC43599.1; -;  
 DR PIR; JC6039; JC6039.  
 DR StyGene; SGI0608; CSGA.  
 KW Fimbria; Signal; Complete proteome.  
 FT SIGNAL 1 20  
 FT CHAIN 21 151 MAJOR CURLIN SUBUNIT.  
 FT CONFLICT 134 151 SYNVROGFGNNATANYQ -> DSYTQVAS (IN  
 REF. 6).  
 FT SEQUENCE 151 AA; 15305 MW; B7DAC0D16B621359 CRC64;  
 Query Match 88.5%; Score 692; DB 1; Length 151;  
 Best Local Similarity 90.7%; Pred. No. 3.2e-51;  
 Matches 137; Conservative 2; Mismatches 12; Indels 0; Gaps 0;  
 QY 1 MKLLKVAAPAAIIVVSGSALAGVVPQWGGGNGNGGNSGSPDSTLSIYQGSANAALQDQ 60  
 Db 1 MKLLKVAAPAAIIVVSGSALAGVVPQWGGGNGNGGNSGSPDSTLSIYQGSANAALQ 60  
 QY 61 LVTRVTVTHMAHAGYNGGADVGCGADNSTIELTQNGFRNNALIDOWNAKNSDITVQYGG 120  
 Db 61 SDARKSETTITQSGYNGGADVGCGADNSTIELTQNGFRNNALIDOWNAKNSDITVQYGG 120  
 QY 121 NNAALVNQATSDSSVMVROVGFNNATANYQ 151  
 Db 121 NNAALVNQATSDSSVMVROVGFNNATANYQ 151

RESULT 2  
 CSGA\_ECOLI STANDARD; PRT; 151 AA.  
 ID CSGA\_ECOLI  
 AC P28307;  
 DT 01-DEC-1992 (Rel. 24, Created)  
 DT 01-OCT-1996 (Rel. 34, Last sequence update)  
 DT 28-FEB-2003 (Rel. 41, Last annotation update)  
 DE Major curlin subunit precursor.  
 GN CSGA OR B1042.  
 OS *Escherichia coli*.  
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;  
 OC Enterobacteriaceae; *Escherichia*.  
 OX NCBI\_TaxID=562;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=K12 / W3110;  
 RX MEDLINE=93211294; PubMed=8459772;  
 RA Olsen A., Arngvist A.;  
 RT "The Rpos sigma factor relieves H-NS-mediated transcriptional  
 repression of *csGA*, the subunit gene of fibronectin-binding curli in  
*Escherichia coli*.";  
 RT Mol. Microbiol. 7:523-536(1993).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=K12 / MC4100;  
 RX MEDLINE=96414468; PubMed=8817489;  
 RA Hammar M., Arngvist A., Bian Z., Olsen A., Normark S.;  
 RT "Expression of two *csg* operons is required for production of  
 fibronectin- and congo red-binding curli polymers in *Escherichia coli*  
 K-12.";  
 RT Mol. Microbiol. 18:661-670(1995).  
 RN [3]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=K12 / MG1655;  
 RX MEDLINE=97426617; PubMed=9278503;  
 RA Blatter F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,  
 Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,  
 Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,  
 Mau B., Shao Y.;  
 RA "The complete genome sequence of *Escherichia coli* K-12.";  
 RT Science 277:1453-1474(1997).  
 RN [4]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=K12;  
 RX MEDLINE=97061202; PubMed=8905232;  
 RA Oshima T., Aiba H., Baba T., Fujita K., Hayaishi K., Honjo A.,  
 Ikemoto K., Inada T., Itoh T., Kajihara M., Kanai K., Kashimoto K.,  
 Kimura S., Kitagawa M., Makino K., Masuda S., Miki T., Mizobuchi K.,  
 Mori H., Motomura K., Nakamura Y., Nishimoto H., Nishio Y., Saito N.,  
 Sampei G., Seki Y., Tagami H., Takemoto K., Wada C., Yamamoto Y.,  
 Yano W., Horiuchi T.;  
 RA "A 718-Kb DNA sequence of the *Escherichia coli* K-12 genome  
 corresponding to the 12.7-28.0 min region on the linkage map.";  
 RT DNA Res. 3:137-155(1996).  
 RN [5]  
 RP SEQUENCE OF 21-40.  
 RC STRAIN=K12 / YNEL;  
 RX MEDLINE=93033873; PubMed=1357528;  
 RA Arngvist A., Olsen A., Pfeifer J., Russell D.G., Normark S.;  
 RT "The Crl protein activates cryptic genes for curli formation and  
 fibronectin binding in *Escherichia coli* HB101.";  
 RT Mol. Microbiol. 6:2443-2452(1992).  
 RN [6]  
 RP SEQUENCE OF 21-31.  
 RX MEDLINE=91310586; PubMed=1677357;  
 RA Collinson S.K., Emeody L., Trust T.J., Kay W.W.;  
 RT "Purification and characterization of thin, aggregative fimbriae from  
*Salmonella enteritidis*.";  
 RL J. Bacteriol. 173:4773-4781(1991).  
 CC -1- FUNCTION: CURLIN IS THE STRUCTURAL SUBUNIT OF THE CURLI. CURLI ARE  
 COILED SURFACE STRUCTURES THAT ASSEMBLE PREFERENTIALLY AT GROWTH  
 TEMPERATURES BELOW 37 DEGREES CELSIUS. CURLI CAN BIND TO  
 FIBRONECTIN.  
 CC -----



RT K-12.";  
 RL Mol. Microbiol. 18:661-670(1995).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=K12 / MG1655;  
 RX MEDLINE=97426617; PubMed=9278503;  
 RA Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,  
 RA Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,  
 RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,  
 RA Mau B., Shao Y.,  
 RA "The complete genome sequence of *Escherichia coli* K-12.";  
 RL Science 277:1453-1474 (1997).  
 RN [3]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=K12;  
 RX MEDLINE=97061202; PubMed=8905232;  
 RA Oshima T., Aiba H., Baba T., Fujita K., Hayashi K., Honjo A.,  
 RA Ikemoto K., Inada T., Itoh T., Kajihara M., Kanai K., Kashimoto K.,  
 RA Kimura S., Kitagawa M., Makino K., Masuda S., Miki T., Mizobuchi K.,  
 RA Mori H., Motomura K., Nakamura Y., Nishimoto H., Nishio Y., Saito N.,  
 RA Sampei G., Seki Y., Tagami H., Takemoto K., Wada C., Yamamoto Y.,  
 RA Yanai M., Horiuchi T.,  
 RA "A 718-kb DNA sequence of the *Escherichia coli* K-12 genome  
 corresponding to the 12.7-28.0 min region on the linkage map.";  
 RL DNA Res. 3:137-155(1996).  
 RN [4]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=O157:H7 / EDL933 / ATCC 700927;  
 RX MEDLINE=21074935; PubMed=11206551;  
 RA Perna N.T., Plunkett G. III, Burland V., Mau B., Glasner J.D.,  
 RA Rose D.J., Mayhew G.F., Evans P.S., Gregor J., Kirkpatrick H.A.,  
 RA Posfai G., Hackett J., Klink S., Boutin A., Shao Y., Miller L.,  
 RA Grobeck E.J., Davis N.W., Lim A., Dimalanta E.T., Potamousis K.,  
 RA Apodaca J., Anantharaman T.S., Lin J., Yen G., Schwartz D.C.,  
 RA Welch R.A., Blattner F.R.,  
 RA "Genome sequence of enterohaemorrhagic *Escherichia coli* O157:H7.";  
 RL Nature 409:529-533(2001).  
 RN [5]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=O157:H7 / RIMD 0509952;  
 RX MEDLINE=21156231; PubMed=11258796;  
 RA Hayashi T., Makino K., Ohnishi M., Kurokawa K., Ishii K., Yokoyama K.,  
 RA Han C.-G., Ohtsubo E., Nakayama K., Murata T., Tanaka M., Tobe T.,  
 RA Iida T., Takami H., Honda T., Sasakawa C., Ogasawara N., Yasunaga T.,  
 RA Kunara S., Shiba T., Hattori M., Shinagawa H.,  
 RA "Complete genome sequence of enterohemorrhagic *Escherichia coli*  
 O157:H7 and genomic comparison with a laboratory strain K-12.";  
 RL DNA Res. 8:111-22(2001).  
 RN [6]  
 RP SEQUENCE OF 1-21 FROM N.A.  
 RC STRAIN=K12;  
 RX MEDLINE=95157246; PubMed=7854117;  
 RA Arngvist A., Olsen A., Normark S.,  
 RA "Sigma S-dependent growth-phase induction of the *csfBA* promoter in  
*Escherichia coli* can be achieved in vivo by sigma 70 in the absence  
 of the nucleoid-associated protein H-NS.";  
 RL Mol. Microbiol. 13:1021-1032(1994).  
 CC -1- FUNCTION: CURLIN IS THE STRUCTURAL SUBUNIT OF THE CURLI. CURLI ARE  
 COILED SURFACE STRUCTURES THAT ASSEMBLE PREFERENTIALLY AT GROWTH  
 TEMPERATURES BELOW 37 DEGREES CELSIUS. CURLI CAN BIND TO  
 FIBRONECTIN. THE MINOR SUBUNIT IS THE NUCLEATION COMPONENT OF  
 CURLIN MONOMERS.  
 CC -1- SIMILARITY: BELONGS TO THE CSGB/CSGB FAMILY.  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 the European Bioinformatics Institute. There are no restrictions on its  
 use by non-profit institutions as long as its content is in no way  
 modified and this statement is not removed. Usage by and for commercial  
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC EMBL; X30754; CAA62281.1; -

DR EMBL; AE000205; AAC74125.1; -  
 DR EMBL; D90741; BAA35831.1; -  
 DR EMBL; AE005315; AAG55787.1; -  
 DR EMBL; AP002554; BAB34842.1; -  
 DR PIR; C90806; C90806.  
 DR PIR; G85665; G85665.  
 DR PIR; S70787; S70787.  
 KW EcoGene; EGI2621; csfB.  
 KW Fimbria; Signal; Complete proteome.  
 FT SIGNAL 1 21 POTENTIAL.  
 FT CHAIN 22 151 MINOR CURLIN SUBUNIT.  
 SQ SEQUENCE 151 AA; 15882 MW; B18D268B96401488 CRC64;  
 Query Match 14.3%; Score 112; DB 1; Length 151;  
 Best Local Similarity 29.8%; Pred. No. 0.0083;  
 Matches 34; Conservative 17; Mismatches 49; Indels 14; Gaps 3;  
 QY 51 GSAAALYDQLVT--RVVTHEAHAGYGVNGADYVGQAGDNSTIELTNGFR----- 98  
 DB 17 GIAAAGYDLANSEYFPAVNELSKSFNQALIIQGAGTWSAQLRGGSKLLAVVAQES 76  
 QY 99 -NNATIDQWNAKSDITVGGYGGNNAALVNVQATSDSSVMVRVQVFGNNATANYQ 151  
 DB 77 SNRAKIDQTGDYNL-AVIDQAGSANDASISQAGYNTAMTIQKSGNKANITQY 129  
 RESULT 5  
 ICEA PANAN STANDARD; PRT; 1322 AA.  
 AC P20469;  
 DT 01-FEB-1991 (Rel. 17, Created)  
 DT 01-FEB-1991 (Rel. 17, Last sequence update)  
 DT 16-OCT-2001 (Rel. 40, Last annotation update)  
 DE Ice nucleation protein InaA.  
 GN INAA.  
 OS Pantoea ananas (Erwinia uredovora).  
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;  
 OC Enterobacteriaceae; Pantoea.  
 OX NCBI\_TaxID=553;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=30092494; PubMed=2599095;  
 RA Abe K., Watabe S., Emori Y., Watanabe M., Arai S.,  
 RA "An ice nucleation active gene of *Erwinia ananas*. Sequence similarity  
 to those of *Pseudomonas* species and regions required for ice  
 nucleation activity.";  
 RT FEMS Lett. 258:297-300(1989).  
 CC -1- FUNCTION: Ice nucleation proteins enable bacteria to nucleate  
 crystallization in supercooled water.  
 CC -1- SUBCELLULAR LOCATION: Outer membrane (By similarity).  
 CC -1- DOMAIN: CONTAINS MANY IMPERFECT REPEATS OF THE CONSENSUS  
 OCTAPEPTIDE A-G-Y-G-S-T-X-T; FURTHER ON A 16-RESIDUE AND A  
 REGIONAL 48-RESIDUE PERIODICITY IS SUPERIMPOSED.  
 CC -1- SIMILARITY: Belongs to the bacterial ice nucleation protein  
 family.  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 the European Bioinformatics Institute. There are no restrictions on its  
 use by non-profit institutions as long as its content is in no way  
 modified and this statement is not removed. Usage by and for commercial  
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC EMBL; X17316; CAA35194.1; -  
 DR PIR; S07053; S07053.  
 DR HSP; P06620; IINA.  
 DR InterPro; IPR000258; Ice\_nucleatn.  
 DR Pfam; PF00818; Ice\_nucleation; 69.  
 DR PRINTS; PR00327; ICNUCLEAIN.  
 DR PROSITE; PS00314; ICE\_NUCLEATION; 49.  
 KW Ice nucleation; Repeat; Outer membrane.  
 FT DOMAIN 162 1281 OCTAPEPTIDE PERIODICITY.

SQ SEQUENCE 1322 AA; 131094 MW; 89B0BE24AA837039 CRC64;  
 Query Match 13.7%; Score 107.5; DB 1; Length 1322;  
 Best Local Similarity 29.9%; Pred. No. 0.22;  
 Matches 38; Conservative 19; Mismatches 49; Gaps 8;  
 QY 34 GGNSSGPGSTLSIYQYGSANAALYDQVTRVVTHEMAHAGYNGADVQCGADNSTIELT 93  
 DB 933 GSTAGPSSSL-AGYGSQTAGYNSILT-----AGYGS-----T 967  
 QY 94 QNGFRNATIDOWNAKNSDITVQYQY-----GNNAALV-----NOTASDSSVMVRQVGF 143  
 DB 968 QTC-----QENSDDTTC-YGSTAGYESSLIAGYSGTQASPKSLM--AGY 1013  
 QY 144 NNATANQ 150  
 DB 1014 SSOATRE 1020

RESULT 6  
 CSGB\_SALTY STANDARD; PRT; 151 AA.  
 AC Q8Z7M3;  
 DT 28-FEB-2003 (Rel. 41, Created)  
 DT 28-FEB-2003 (Rel. 41, Last sequence update)  
 DT 10-OCT-2003 (Rel. 42, Last annotation update)  
 DE Minor curlin subunit precursor.  
 GN CSGB OR STY1180 OR T1777.  
 OS Salmonella typhi.  
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;  
 OC Enterobacteriaceae; Salmonella.  
 OX NCBI\_TaxID=601;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=CT18;  
 RX MEDLINE=21534947; PubMed=11677608;  
 RA Parkhill J., Dougan G., James K.D., Thomson N.R., Pickard D., Wain J.,  
 RA Churcher C., Mungall K.L., Bentley S.D., Holden M.T.G., Sebahia M.,  
 RA Baker S., Basham P., Brooks K., Chillingworth T., Connerton P.,  
 RA Cronin A., Davies P., Davies R.M., Dowd L., White N., Farrar J.,  
 RA Feltwell T., Hamlin N., Haque A., Hien T.T., Holtroyd S., Jagels K.,  
 RA Krogil M.A., Larsen T.S., Leather S., Moule S., O'Gaora P., Parry C.,  
 RA Quail M.A., Rutherford K., Simmonds M., Skelton J., Stevens K.,  
 RA Whitehead S., Barrett B.G.;  
 RT "Complete genome sequence of a multiple drug resistant Salmonella  
 RT enterica serovar Typhi CT18.";  
 RL Nature 413:848-852(2001).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=TY2 / ATCC 700931;  
 RX MEDLINE=22531367; PubMed=12644504;  
 RA Deng W., Liou S.-R., Plunkett G. III, Mayhew G.F., Rose D.J.,  
 RA Burland V., Kodyanni V., Schwartz D.C., Blattner F.R.;  
 RT "Comparative genomics of Salmonella enterica serovar Typhi strains Ty2  
 RT and CT18.";  
 RL J. Bacteriol. 185:2330-2337(2003).  
 CC -!- FUNCTION: CURLIN IS THE STRUCTURAL SUBUNIT OF THE CURLI. CURLI ARE  
 CC COILED SURFACE STRUCTURES THAT ASSEMBLE PREFERENTIALLY AT GROWTH  
 CC TEMPERATURES BELOW 37 DEGREES CELSIUS. CURLI CAN BIND TO  
 CC FIBRONECTIN. THE MINOR SUBUNIT IS THE NUCLEATION COMPONENT OF  
 CC CURLIN MONOMERS.  
 CC -!- SIMILARITY: BELONGS TO THE CSGB/CSGB FAMILY.  
 CC  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 DR EMBL; AL627269; CAD08267.1; -  
 DR EMBL; AE016840; AAO69400.1; -

KW Fimbria; Signal; Complete proteome.  
 FT SIGNAL 1 21 POTENTIAL.  
 FT CHAIN 22 151 MINOR CURLIN SUBUNIT.  
 SQ SEQUENCE 151 AA; 16254 MW; 161C54326E573495 CRC64;  
 Query Match 13.7%; Score 107; DB 1; Length 151;  
 Best Local Similarity 30.4%; Pred. No. 0.02;  
 Matches 35; Conservative 17; Mismatches 47; Indels 16; Gaps 5;  
 QY 51 GSANAALYD--QLVTRVVTHEMAHAGYNGADVQCGADNST-----IELTQNGF 97  
 DB 17 GIATATNYDLARSEVNFVAVNELSKSSFNQAAIIQVGTDSARVVRQEGSKLLSVISQEGE 76  
 QY 98 RNNATIDOWNAKNSDIT-VQYQYGGNNAALVNOTASDSSVMVRQVGFQGNNTANQY 151  
 DB 77 NNRAKVDQ--AGNYNFAYIEQTGNANDASISQSAIGNSAAIIQKSGNKANITQY 129  
 RESULT 7  
 CSGB\_SALTY STANDARD; PRT; 151 AA.  
 AC P55226;  
 DT 01-OCT-1996 (Rel. 34, Created)  
 DT 01-OCT-1996 (Rel. 34, Last sequence update)  
 DT 28-FEB-2003 (Rel. 41, Last annotation update)  
 DE Minor curlin subunit precursor (Fimbrin SEF17 minor subunit).  
 GN CSGB OR AGFB OR STM1143.  
 OS Salmonella typhimurium, and  
 OS Salmonella enteritidis.  
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;  
 OC Enterobacteriaceae; Salmonella.  
 OX NCBI\_TaxID=602, 592;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC SPECIES=S.typhimurium; STRAIN=SR-11;  
 RX MEDLINE=98117058; PubMed=9457880;  
 RA Romling U., Bian Z., Hammar M., Sierzalta W.D., Normark S.;  
 RT "Curli fibers are highly conserved between Salmonella typhimurium and  
 RT Escherichia coli with respect to operon structure and regulation.";  
 RL J. Bacteriol. 180:722-731(1998).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC SPECIES=S.typhimurium; STRAIN=LT2 / SGSC1412 / ATCC 700720;  
 RX MEDLINE=21534948; PubMed=11677609;  
 RA McClelland M., Sanderson K.E., Spieth J., Clifton S.W., Latreille P.,  
 RA Courtney L., Porwollik S., Ali J., Dante M., Du F., Hou S., Layman D.,  
 RA Leonard S., Nguyen C., Scott K., Holmes A., Grewal N., Mulvaney E.,  
 RA Ryan E., Sun H., Florea L., Miller W., Stoneking T., Nhan M.,  
 RA Waterston R., Wilson R.K.;  
 RT "Complete genome sequence of Salmonella enterica serovar Typhimurium  
 RT LT2.";  
 RL Nature 413:852-856(2001).  
 RN [3]  
 RP SEQUENCE FROM N.A.  
 RC SPECIES=S.enteritidis; STRAIN=27655-3B;  
 RX MEDLINE=96144512; PubMed=8550497;  
 RA Collinson S.K., Clouthier S.C., Doran J.L., Baner P.A., Kay W.W.;  
 RT "Salmonella enteritidis agfBAC operon encoding thin, aggregative  
 RT fimbriae.";  
 RL J. Bacteriol. 178:662-667(1996).  
 CC -!- FUNCTION: CURLIN IS THE STRUCTURAL SUBUNIT OF THE CURLI. CURLI ARE  
 CC COILED SURFACE STRUCTURES THAT ASSEMBLE PREFERENTIALLY AT GROWTH  
 CC TEMPERATURES BELOW 37 DEGREES CELSIUS. CURLI CAN BIND TO  
 CC FIBRONECTIN. THE MINOR SUBUNIT IS THE NUCLEATION COMPONENT OF  
 CC CURLIN MONOMERS.  
 CC -!- SIMILARITY: BELONGS TO THE CSGB/CSGB FAMILY.  
 CC  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC EMBL; AJ002301; CAA05316.1; --  
 DR EMBL; AE008749; AAL20073.1; --  
 DR EMBL; U43280; AAC43598.1; --  
 DR PIR; JC6040; JC6040.  
 DR StyGene; SGI0609; csGB.  
 KW Fimbrria; Signal; Complete proteome.  
 FT SIGNAL 1 21 POTENTIAL.  
 FT CHAIN 22 151 MINOR CURLIN SUBUNIT.  
 SQ SEQUENCE 151 AA; 16182 MW; C0FC5430E6DD361D CRC64;

Query Match 13.7%; Score 107; DB 1; Length 151;  
 Best Local Similarity 30.4%; Pred. No. 0.022;  
 Matches 35; Conservative 17; Mismatches 47; Indels 16; Gaps 5;

QY 51 GSANAALYD--QLVTRVWTHMAHAGYNGADVGQ--GADNST-----IELTQNGF 97  
 DB 17 GIATNYDLARSEYNFAVNEKSSFNQAAIIIGVGTNSARVRQEGSKLLSVISQEGG 76

QY 98 RNNATIDWNKNSDIT--VGQYGGNNAALVNQTSDDSSVMVRQVGFNNATNQY 151  
 DB 77 NNRKAVDQ--AGNYEAFYIEQTGNANDASISOSAYGNSAAIIQKSGNKNATIQY 129

RESULT 8

ICEN PANAN STANDARD; PRT; 1034 AA.

ID ICEN PANAN  
 AC Q47879;  
 DT 01-NOV-1997 (Rel. 35, Created)  
 DT 01-NOV-1997 (Rel. 35, Last sequence update)  
 DT 16-OCT-2001 (Rel. 40, Last annotation update)  
 DE Ice nucleation protein InaU.  
 GN INAU.  
 OS Pantoea ananas (Erwinia uredovora).  
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;  
 OC Enterobacteriaceae; Pantoea.  
 OX NCBI\_TaxID=553;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=KUI-3;  
 RX MEDLINE=94264407; PubMed=7764866;  
 RA Michigami Y., Watabe S., Abe K., Obata H., Arai S.;  
 RT "Cloning and sequencing of an ice nucleation active gene of Erwinia uredovora.";  
 RL Biosci. Biotechnol. Biochem. 58:762-764 (1994).  
 CC -!- FUNCTION: Ice nucleation proteins enable bacteria to nucleate  
 CC crystallization in supercooled water.  
 CC -!- SUBCELLULAR LOCATION: Outer membrane.  
 CC -!- DOMAIN: CONTAINS IMPERFECT REPEATS OF A CONSENSUS OCTAPEPTIDE  
 CC A-G-Y-G-S-T-X-T; FURTHER ON A 16-RESIDUE AND A REGIONAL 48-RESIDUE  
 CC PERIODICITY IS SUPERIMPOSED.  
 CC -!- MISCELLANEOUS: A STRUCTURAL MODEL IS SUGGESTED IN WHICH THE ICE  
 CC NUCLEATION PROTEIN DISPLAYS A SYMMETRY RELATED TO THAT OF ICE.  
 CC -!- SIMILARITY: Belongs to the bacterial ice nucleation protein  
 CC family.

-----  
 This SWISS-PROT entry is copyright. It is produced through a collaboration  
 between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 the European Bioinformatics Institute. There are no restrictions on its  
 use by non-profit institutions as long as its content is in no way  
 modified and this statement is not removed. Usage by and for commercial  
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

-----  
 EMBL; D14992; BAA03636.1; --  
 DR PIR; JC2143; JC2143.  
 DR HSP; P06620; IINA.  
 DR InterPro; IPR000258; Ice\_nucleatn.  
 DR Pfam; PF00818; Ice\_nucleatn.  
 DR PRINTS; PRO0327; ICENUCLEATN.  
 DR PROSITE; PS00314; ICE\_NUCLEATION; 51.  
 KW Ice nucleation; Repeat; Outer membrane.

FT DOMAIN 162 993 OCTAPEPTIDE PERIODICITY  
 SQ SEQUENCE 1034 AA; 103378 MW; FA222523D333EADD CRC64;

Query Match 13.6%; Score 106; DB 1; Length 1034;  
 Best Local Similarity 29.9%; Pred. No. 0.22;  
 Matches 40; Conservative 19; Mismatches 37; Indels 38; Gaps 9;

QY 27 GGGNHNCGGSSGPDSTLSIYQYGSANAALYDQLVTRVWTHMAHAGYNGADVGQGD 86  
 DB 531 GYGSTSTAGANSS-----LIAGYGTQTSYNSVL-----AGYGTQTAREGSD 575

QY 87 NSTIELFQNGFRNNATIDWNKNSDITVGQY-----GNNAALV-----NOTASDSSVM 136  
 DB 576 -----LTA-GYGSTQTAQE-----NSDLTTG-YGSTAGYDSSLIAGYGTQTAQYHSIL 624

QY 137 VRQVGFNNATNQ 150  
 DB 625 T--AGYGTQTAQE 636

RESULT 9

ICEN ERWHE STANDARD; PRT; 1258 AA.

ID ICEN ERWHE  
 AC P16239;  
 DT 01-APR-1990 (Rel. 14, Created)  
 DT 01-APR-1990 (Rel. 14, Last sequence update)  
 DT 16-OCT-2001 (Rel. 40, Last annotation update)  
 DE Ice nucleation protein.  
 GN ICEE.  
 OS Erwinia herbicola.  
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;  
 OC Enterobacteriaceae; Pantoea.  
 OX NCBI\_TaxID=549;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=M1;  
 RX MEDLINE=90152370; PubMed=2515997;  
 RA Warren G.J., Corotto L.V.;  
 RT "The consensus sequence of ice nucleation proteins from Erwinia herbicola, Pseudomonas fluorescens and Pseudomonas syringae.";  
 RL Gene 85:239-242 (1989).  
 CC -!- FUNCTION: Ice nucleation proteins enable bacteria to nucleate  
 CC crystallization in supercooled water.  
 CC -!- SUBCELLULAR LOCATION: Outer membrane.  
 CC -!- DOMAIN: CONTAINS 126 IMPERFECT REPEATS OF A CONSENSUS OCTAPEPTIDE  
 CC A-G-Y-G-S-T-X-T; FURTHER ON A 16-RESIDUE AND A REGIONAL 48-RESIDUE  
 CC PERIODICITY IS SUPERIMPOSED.  
 CC -!- MISCELLANEOUS: A STRUCTURAL MODEL IS SUGGESTED IN WHICH THE ICE  
 CC NUCLEATION PROTEIN DISPLAYS A SYMMETRY RELATED TO THAT OF ICE.  
 CC -!- SIMILARITY: Belongs to the bacterial ice nucleation protein  
 CC family.

-----  
 This SWISS-PROT entry is copyright. It is produced through a collaboration  
 between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 the European Bioinformatics Institute. There are no restrictions on its  
 use by non-profit institutions as long as its content is in no way  
 modified and this statement is not removed. Usage by and for commercial  
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

-----  
 EMBL; M26382; AAA24823.1; --  
 DR PIR; JQ0188; JQ0188.  
 DR HSP; P06620; IINA.  
 DR InterPro; IPR000258; Ice\_nucleatn.  
 DR Pfam; PF00818; Ice\_nucleatn.  
 DR PRINTS; PRO0327; ICENUCLEATN.  
 DR PROSITE; PS00314; ICE\_NUCLEATION; 45.  
 KW Ice nucleation; Repeat; Outer membrane.  
 FT DOMAIN 162 1217 OCTAPEPTIDE PERIODICITY.  
 SQ SEQUENCE 1258 AA; 125084 MW; 590E8A13007FBD4 CRC64;

Query Match 13.4%; Score 104.5; DB 1; Length 1258;  
 Best Local Similarity 29.5%; Pred. No. 0.35;





PFAM: PF03797; Autotransporter; 1.  
DR TIGR01414; autotrans. barl; 2.  
KW Antigen; S-layer; Cell wall; Complete proteome.  
FT CHAIN 1335 1655 120 KDa SURFACE-EXPOSED PROTEIN.  
FT CHAIN 1335 1655 32 KDa BETA PEPTIDE.  
FT VARIANT 61 61 P -> A (IN STRAIN INDIAN TICK TYPHUS).  
FT VARIANT 75 75 G -> S (IN STRAIN INDIAN TICK TYPHUS).  
FT VARIANT 78 78 K -> N (IN STRAIN INDIAN TICK TYPHUS).  
FT VARIANT 251 251 V -> A (IN STRAIN INDIAN TICK TYPHUS).  
FT VARIANT 413 413 N -> D (IN STRAIN INDIAN TICK TYPHUS).  
FT VARIANT 499 499 I -> V (IN STRAIN INDIAN TICK TYPHUS).  
FT VARIANT 959 959 A -> T (IN STRAIN INDIAN TICK TYPHUS).  
FT VARIANT 988 988 R -> L (IN STRAIN INDIAN TICK TYPHUS).  
FT VARIANT 1139 1139 R -> L (IN STRAIN INDIAN TICK TYPHUS).  
FT VARIANT 1139 1139 KD -> GH (IN REF. 3).  
FT CONFLICT 353 354 F -> S (IN REF. 3).  
FT CONFLICT 776 776 E -> D (IN REF. 3).  
FT CONFLICT 1159 1159 G -> S (IN REF. 3).  
FT CONFLICT 1177 1177 H -> R (IN REF. 3).  
FT CONFLICT 1492 1492 H -> R (IN REF. 3).  
SQ SEQUENCE 1655 AA; 168342 MW; 494919377D5FCE37 CRC64;

Query Match 13.1%; Score 102.5; DB 1; Length 1655;  
Best Local Similarity 26.8%; Pred. No. 0.73;  
Matches 39; Conservative 18; Mismatches 55; Indels 31; Gaps 6;

QY 28 GGNHNG--GGSSGGPDSLTISYQGSANAALYDQIVTR-----VVTHEMAHAGYNGGA 79  
DB 1211 GGARFNGTLGGPNFVVTGNSRFRVNYGLIARAAGDYVITRTNAENVVTINDIANSFGGAP 1270  
QY 80 DVGCGADNSTIETQNGFRNATIDQWNAKNSDITVGYCGNNAALVNCATSDSSVMVR- 138  
DB 1271 GVGQNV-TTFVNAITAAVNNLL-----AKNS-----ANSFVGAIVDTSAIIN 1317  
QY 139 -----QVFGNNATANYQY 151  
DB 1318 AQLDVAKDIQALGNRLGALRY 1339

RESULT 12  
ID\_OVO\_DROME STANDARD; PRT; 1028 AA.  
AC P51521; Q9XZU4;  
DT 01-OCT-1996 (Rel. 34, Created)  
DT 01-OCT-1996 (Rel. 34, Last sequence update)  
DT 10-OCT-2003 (Rel. 42, Last annotation update)  
DE OVO protein (Shaven baby protein).  
GN OVO OR SVE.  
OS Drosophila melanogaster (Fruit fly).  
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;  
OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;  
OC Ephydroidea; Drosophilidae; Drosophila.  
OX NCBI\_TaxID=7227;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC TISSUE=Ovary;  
RX MEDLINE=95021209; PubMed=7935398;  
RA Garfinkel M.D., Wang J., Liang Y., Mahowald A.P.;  
RT "Multiple products from the shavenbaby-ovo gene region of Drosophila melanogaster: relationship to genetic complexity.";  
RL Mol. Cell. Biol. 14:6809-6818(1994).  
RN [2]  
RP SEQUENCE FROM N.A.  
RC STRAIN=Oregon-R;  
RX MEDLINE=91293102; PubMed=1712294;  
RA Mevel-Ninio M.T.M., Terracol R., Kafatos F.C.;  
RT "The ovo gene of Drosophila encodes a zinc finger protein required for female germ line development.";  
RL EMO J. 10:2259-2266(1991).  
CC -!- FUNCTION: REQUIRED FOR SURVIVAL AND DIFFERENTIATION OF FEMALE GERM LINE CELLS. PLAYS A ROLE IN GERM LINE SEX DETERMINATION.  
CC -!- SUBCELLULAR LOCATION: Nuclear (Potential).  
CC -!- DEVELOPMENTAL STAGE: FIRST APPEARS IN THE GERMARIUM AND ACCUMULATES IN NURSE CELLS DURING OOGENESIS. STORED IN THE EGG, BUT IS RAPIDLY LOST IN THE EMBRYOS EXCEPT FOR ITS CONTINUED

PRESENCE IN THE GERM LINE PRECURSOR POLE CELLS.  
-!- SIMILARITY: Contains 4 C2H2-type zinc fingers.  
-----  
This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See http://www.isb-sib.ch/announce/ or send an email to license@isb-sib.ch).  
-----  
EMBL: U11383; AAB60216.1; --  
EMBL: X59772; CAB36921.1; ALT\_SEQ.  
PIR: A56038; A56038.  
HSSP: P07248; 2ADR.  
DR TRANSFAC; T00669; --  
DR FlyBase; FBgn0003028; ovo.  
DR InterPro; IPR007087; Znf C2H2.  
DR Pfam; PF00096; zf-C2H2; 3.  
DR SMART; SM00355; Znf C2H2; 4.  
DR PROSITE; PS00028; ZINC\_FINGER\_C2H2\_1; 3.  
DR PROSITE; PS00157; ZINC\_FINGER\_C2H2\_2; 3.  
KW Zinc-finger; Metal-binding; DNA-binding; Repeat; Nuclear protein;  
KW Transcription regulation.  
FT DOMAIN 62 66 POLY-ALA.  
FT DOMAIN 72 77 POLY-GLY.  
FT DOMAIN 80 85 POLY-GLY.  
FT DOMAIN 98 108 POLY-GLY.  
FT DOMAIN 144 152 POLY-HIS.  
FT DOMAIN 153 159 POLY-ASN.  
FT DOMAIN 336 339 POLY-GLN.  
FT DOMAIN 347 353 POLY-GLN.  
FT DOMAIN 357 361 POLY-GLN.  
FT DOMAIN 410 414 POLY-GLN.  
FT DOMAIN 418 422 POLY-GLN.  
FT DOMAIN 426 432 POLY-GLN.  
FT DOMAIN 445 453 POLY-GLN.  
FT DOMAIN 456 459 POLY-GLN.  
FT DOMAIN 466 474 POLY-GLN.  
FT DOMAIN 497 517 POLY-ALA.  
FT DOMAIN 524 529 POLY-SER.  
FT DOMAIN 549 558 POLY-ALA.  
FT DOMAIN 639 651 POLY-ALA.  
FT DOMAIN 717 725 POLY-ALA.  
FT DOMAIN 797 802 POLY-GLN.  
FT DOMAIN 820 823 POLY-GLN.  
FT DOMAIN 826 832 POLY-GLN.  
FT DOMAIN 874 896 C2H2-TYPE 1.  
FT ZN\_FING 902 924 C2H2-TYPE 2.  
FT ZN\_FING 930 953 C2H2-TYPE 3.  
FT ZN\_FING 969 992 C2H2-TYPE 4.  
FT CONFLICT 647 647  
SQ SEQUENCE 1028 AA; 110620 MW; D7068BB2BC0F677 CRC64;

Query Match 12.5%; Score 97.5; DB 1; Length 1028;  
Best Local Similarity 28.7%; Pred. No. 1.1;  
Matches 43; Conservative 12; Mismatches 60; Indels 35; Gaps 7;

QY 3 LLKVAFAAIVVSGSALAGVVPQGGGNGHNGSSGSPDSTLSIYQGSANAALYDQIV 62  
DB 59 LQNAAAAAVIMSAGS-----GGCTGNGGCGAGPGGSPSANGSGGGGG----- 104  
QY 63 TRVTHEMAHAGYNGADVGOGADNSTIETQNGFRNATIDQWNAKNSDI-----TVGQY 118  
DB 105 -----GGGYINCGGVG-GPNN-----LDGNLLNFASVSYNYSNFKFHHHHHHQH 152  
QY 119 GGNNAALVNQATSDSSVMVRQVGFNNATA 148  
DB 153 NNNN-----NNNGGQTSMMGHFF-YGGNPSA 177

RESULT 13  
ICEV\_PSEX

```

ID ICEV_PSEX STANDARD; PRT; 1196 AA.
AC Q3479;
DT 15-DEC-1998 (Rel. 37, Created)
DT 15-DEC-1998 (Rel. 37, Last sequence update)
DE 16-OCT-2001 (Rel. 40, Last annotation update)
DE Ice nucleation protein.
GN INAV
OS Pseudomonas syringae.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;
OC Pseudomonadaceae; Pseudomonas.
OX NCBI_TaxID=317;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=INA5.
RC MEDLINE=97462815; PubMed=9323042;
RX Schmid D., Pridmore D., Capitani G., Battistutta R., Neeser J.-R.,
RA Jann A.;
RT "Molecular organisation of the ice nucleation protein Inav from
RT Pseudomonas syringae";
RL PEBB Lett. 414:590-594(1997).
CC -!- FUNCTION: Ice nucleation proteins enable bacteria to nucleate
CC crystallization in supercooled water.
CC -!- SUBUNIT: MEMBRANE ENVIRONMENT OR AGGREGATION SEEMS TO BE REQUIRED
CC FOR ICE NUCLEATION ACTIVITY.
CC -!- SUBCELLULAR LOCATION: Outer membrane (By similarity).
CC -!- DOMAIN: CONTAINS MANY IMPERFECT REPEATS OF THE CONSENSUS
CC OCTAPEPTIDE A-G-Y-G-S-T-L-T; FURTHER ON A 16-RESIDUE AND A
CC REGIONAL 48-RESIDUE PERIODICITY IS SUPERIMPOSED.
CC -!- SIMILARITY: Belongs to the bacterial ice nucleation protein
CC family.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; AJ001086; CAA04521.1; -
DR HSP; P06620; 11NA.
DR InterPro; IPR000258; Ice_nucleatn.
DR Pfam; PF00818; Ice_nucleation; 61.
DR PRINTS; PR00327; ICENUCLEATN.
DR PROSITE; PS00314; ICE_NUCLEATION; 42.
KW Ice nucleation; Repeat; Outer membrane.
FT DOMAIN 172 1147 OCTAPEPTIDE PERIODICITY.
SQ SEQUENCE 1196 AA; 117991 MW; C9E9974CB171E68 CRC64;

Query Match 12.3%; Score 96.5; DB 1; Length 1196;
Best Local Similarity 26.7%; Pred. No. 1.6;
Matches 43; Conservative 24; Mismatches 39; Indels 55; Gaps 11;

QY 16 GSALAG-----VYPWGG--GGNH-----GGGSSGPDSTLSIYQYGSANAALYDOL 61
DB 175 GTSLDGNNRLIAGYGSNETAGNSDLIAGYGTGTAGSDSSL-VAGYGTQTAGGDSA 233

QY 62 VTRVTHEMAHAGYGN-----GADVGGADNSTIELTQNGFRNNATIDOWN 107
DB 234 LT-----ACYGSTQTAREGNSLTCYGTGTAGSDSLIA-----GYGSTQT----S 276

QY 108 AKNSDITVQYGGNNAALVNOTASDSSVNVVRQVFGNATA 148
DB 277 GSDSSLTAG-YGS-----TQTAQEGSNLT--AGYGSTGTA 308

RESULT 14
OMP_RICJA
ID_OMP_RICJA STANDARD; PRT; 1656 AA.
AC O06653;
DT 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DE 16-OCT-2001 (Rel. 40, Last annotation update)

DE Outer membrane protein B precursor (168 kDa surface-layer protein)
DE (Surface protein antigen) (Cell surface antigen 5) (Scas) (rOmpB)
DE (rOmp B) (Contains: 120 kDa surface-exposed protein (Surface protein
DE antigen) (120 kDa outer membrane protein ompB); 32 kDa beta peptide).
GN OMPB
OS Rickettsia japonica.
OC Bacteria; Proteobacteria; Alphaproteobacteria; Rickettsiales;
OC Rickettsiaceae; Rickettsiae; Rickettsia.
OX NCBI_TaxID=35790;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=YH;
RC Uchiyama T.;
RT "Sequencing of the gene encoding the protein rOmp B of Rickettsia
RT japonica.";
RL Submitted (MAY-1997) to the EMBL/GenBank/DBJ databases.
CC -!- FUNCTION: THE 120 kDa SURFACE-EXPOSED PROTEIN IS A MAJOR
CC STRUCTURAL PROTEIN WHICH MAY PLAY A ROLE AS A RICKETTSIAL
CC VIRULENCE FACTOR AND/OR IMMUNOGEN DURING INFECTION (By
CC similarity).
CC -!- FUNCTION: THE 32 kDa BETA PEPTIDE MAY SERVE AS A MEMBRANE ANCHOR
CC (By similarity).
CC -!- SUBCELLULAR LOCATION: Cell wall. This bacterium is covered by a s-
CC layer with hexagonal symmetry.
CC -!- SIMILARITY: BELONGS TO THE RICKETTSIAE OMPA/OMPB FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; AB003681; BAA20138.1; -
DR InterPro; IPR006315; Autotransport.
DR InterPro; IPR005546; Autotransporter.
DR Pfam; PF03797; Autotransporter; 1.
DR TIGRFAMs; TIGR01414; autotrans_bar1; 2.
KW Antigen; S-layer; Cell wall.
FT CHAIN 1 1338 120 kDa SURFACE-EXPOSED PROTEIN.
FT CHAIN 1339 1656 32 kDa BETA PEPTIDE.
FT DOMAIN 528 533 POLY-GLY.
SQ SEQUENCE 1656 AA; 168097 MW; 3132A69C9DD5999F CRC64;

Query Match 12.2%; Score 95.5; DB 1; Length 1656;
Best Local Similarity 25.0%; Pred. No. 2.8;
Matches 36; Conservative 21; Mismatches 52; Indels 35; Gaps 7;

QY 28 GGNHNGGSSGPDSTLS-----IYQYGSANAALYDOLVTR-----VVTHEMAHAGYGN 77
DB 1215 GGARFN--GTLGGNFVTGNSRNFVNYGLIRAAQDYVITRTNNAENIVTNDITNSPFG 1272

QY 78 GADVGGADNSTIELTQNGFRNNATIDOWNAKSDITVQYGGNNAALVNOTASDSSVMV 137
DB 1273 APGVGVNV-TTFVNATTAANNLL-----AKNS-----ADSFVGTIVTDSAAI 1319

QY 138 R-----QVGFGNNTANQY 151
DB 1320 TNAQLDVAQDIQAQGLNRLGALRY 1343

RESULT 15
GP63_LEICH
ID_GP63_LEICH STANDARD; PRT; 599 AA.
AC P15706;
DT 01-APR-1990 (Rel. 14, Created)
DT 01-APR-1990 (Rel. 14, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Leishmanolysin precursor (EC 3.4.24.36) (Cell surface protease)
DE (Major surface glycoprotein) (GP63 protein) (Promastigote surface
DE endopeptidase).
GN GP63.

```

OS Leishmania chagasi.  
 CC Eukaryota; Euglenozoa; Kinetoplastida; Trypanosomatidae; Leishmania.  
 OX NCBI\_TaxID=44271;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA MEDLINE=90205976; PubMed=2320059;  
 RA Miller R.A., Reed S.G., Parsons M.;  
 RT "Leishmania gp63 molecule implicated in cellular adhesion lacks an  
 RT Arg-Gly-Asp sequence.";  
 RL Mol. Biochem. Parasitol. 39:267-274 (1990).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RA MEDLINE=92112918; PubMed=1370484;  
 RA Ramamoorthy R., Donelson J.E., Paetz K.E., Maybodi M., Roberts S.C.,  
 RA Wilson M.E.;  
 RT "Three distinct RNAs for the surface protease gp63 are differentially  
 RT expressed during development of Leishmania donovani Chagasi  
 RT promastigotes to an infectious form.";  
 RL J. Biol. Chem. 267:1888-1895 (1992).  
 CC -!- FUNCTION: Has an integral role during the infection of macrophages  
 CC in the mammalian host.  
 CC -!- CATALYTIC ACTIVITY: Preference for hydrophobic residues at P1 and  
 CC P1' and basic residues at P2 and P3'. A model nonapeptide is  
 CC cleaved at Ala-Tyr-[Leu-Lys-Lys-  
 CC -!- COFACTOR: Binds 1 zinc ion per subunit (By similarity).  
 CC -!- SUBCELLULAR LOCATION: Attached to the membrane by a GPI-anchor.  
 CC -!- SIMILARITY: Belongs to peptidase family M8.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@sib-sib.ch](mailto:license@sib-sib.ch)).  
 CC -----  
 CC EMBL; M80672; AAA29238.1; -;  
 CC EMBL; M28527; AAA29235.1; -;  
 CC PIR; A44951; A44951.  
 CC HSP; P08148; 1LML.  
 CC MEROPS; M08.001; -;  
 CC InterPro; IPR006025; Pept\_M\_Zn\_BS.  
 CC InterPro; IPR001577; Peptidase\_M8.  
 CC Pfam; PF01457; Peptidase\_M8; 1.  
 CC PRINTS; PR00782; LSHMANOLYSIN.  
 CC PROSITE; PS00142; ZINC\_PROTEASE; 1.  
 CC Hydrolase; Metalloprotease; Glycoprotein; Metal-binding; Zinc;  
 CC Zymogen; Signal; Cell adhesion; GPI-anchor; Lipoprotein.  
 FT SIGNAL 1 39  
 FT PROPEP 40 97  
 FT CHAIN 98 574  
 FT CHAIN 98 574  
 FT PROPEP 575 599  
 FT METAL 261 261  
 FT ACT SITE 262 262  
 FT METAL 265 265  
 FT METAL 331 331  
 FT DISULFID 122 139  
 FT DISULFID 188 227  
 FT DISULFID 311 383  
 FT DISULFID 390 452  
 FT DISULFID 403 422  
 FT DISULFID 412 486  
 FT DISULFID 463 507  
 FT DISULFID 512 562  
 FT DISULFID 532 555  
 FT CARBOHYD 297 297  
 FT CARBOHYD 394 394  
 FT LIPID 574 574  
 FT SEQUENCE 599 AA; 63848 MW; 746730AE8E2A2E7C CRC64;  
 Query Match 12.0%; Score 94; DB 1; Length 599;  
 Best Local Similarity 54.8%; Pred. No. 1.2;

Matches 23; Conservative 5; Mismatches 8; Indels 6; Gaps 2;  
 QY 52 SAN-AALYDQLVTRVVTHEMAHA-----GYNGADVGGQADN 87  
 DB 244 AANIASRYDQLVTRVVTHEMAHALGFSVGFEGARILESIN 285  
 Search completed: August 2, 2004, 14:49:32  
 Job time : 6.3 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: August 2, 2004, 14:39:17 ; Search time 29.7 Seconds

(without alignments)  
1604.150 Million cell updates/sec

Title: US-09-543-407-26

Perfect score: 782

Sequence: 1 MKLLKVAFAAIVVSGSALA.....DSSVMVRQVGFNNATANQY 151

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1017041 seqs, 3,551,820 residues

Total number of hits satisfying chosen parameters: 1017041

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

SPTREMBL\_25:\*

- 1: sp\_archaea:\*
- 2: sp\_bacteria:\*
- 3: sp\_fungi:\*
- 4: sp\_human:\*
- 5: sp\_invertebrate:\*
- 6: sp\_mammal:\*
- 7: sp\_mmc:\*
- 8: sp\_organelle:\*
- 9: sp\_phage:\*
- 10: sp\_plant:\*
- 11: sp\_rodent:\*
- 12: sp\_virus:\*
- 13: sp\_vertibrate:\*
- 14: sp\_unclassified:\*
- 15: sp\_rvirus:\*
- 16: sp\_bacteriap:\*
- 17: sp\_archaeap:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	683	87.3	152	Q33802	O33802 salmonella
2	580.5	74.2	150	Q7X243	Q7X243 citrobacter
3	532	68.0	149	Q7X240	Q7X240 citrobacter
4	508.5	65.0	152	Q8CW63	Q8CW63 escherichia
5	442.5	56.6	150	Q7X237	Q7X237 enterobacte
6	304	38.9	76	Q54069	Q54069 salmonella
7	122	15.6	29	Q93J35	Q93J35 escherichia
8	121	15.5	502	Q8EIH4	Q8EIH4 shewanella
9	114.5	14.6	145	Q8U6N9	Q8U6N9 agrobacteri
10	113	14.5	2174	Q92U08	Q92U08 rhizobium m
11	112	14.3	151	Q7UCZ1	Q7UCZ1 shigella fl
12	112	14.3	160	Q8CW64	Q8CW64 escherichia
13	112	14.3	160	Q83RU7	Q83RU7 shigella fl
14	110	14.1	151	Q7X244	Q7X244 citrobacter
15	108.5	13.9	133	Q8PD38	Q8PD38 xanthomonas
16	107	13.7	1422	Q8EFU3	Q8EFU3 shewanella

17	105.5	13.5	645	16	Q7U1C5	Q7U1C5 mycobacteri
18	105.5	13.5	646	16	O53818	O53818 mycobacteri
19	105	13.4	589	16	Q7X53	Q7X53 mycobacteri
20	105	13.4	590	16	O53309	O53309 mycobacteri
21	105	13.4	603	16	Q8VU66	Q8VU66 mycobacteri
22	104.5	13.4	151	2	Q7X238	Q7X238 enterobacte
23	104	13.3	490	16	Q8EY99	Q8EY99 leptospira
24	103.5	13.2	91	2	Q9S3J8	Q9S3J8 escherichia
25	103.5	13.2	1306	2	Q93N36	Q93N36 pantoea ana
26	102.5	13.1	552	16	P96840	P96840 mycobacteri
27	102.5	13.1	552	16	Q7TW76	Q7TW76 mycobacteri
28	102.5	13.1	623	16	Q8V1Y0	Q8V1Y0 mycobacteri
29	102.5	13.1	1616	2	Q9KKA1	Q9KKA1 rickettsia
30	102	13.0	130	16	Q8E9J14	Q8E9J14 bradyrhizob
31	102	13.0	139	16	O8EIH3	O8EIH3 shewanella
32	101.5	13.0	152	2	Q7X241	Q7X241 citrobacter
33	101.5	13.0	1616	2	Q9KKA5	Q9KKA5 rickettsia
34	101	12.9	157	16	Q88HG0	Q88HG0 pseudomonas
35	100.5	12.9	1616	2	Q9KKA9	Q9KKA9 rickettsia
36	100	12.8	191	10	Q7XDR3	Q7XDR3 oryza sativ
37	100	12.8	3275	16	Q8VKM3	Q8VKM3 mycobacteri
38	100	12.8	3300	16	O06304	O06304 mycobacteri
39	100	12.8	3507	16	Q7U270	Q7U270 mycobacteri
40	99.5	12.7	1613	2	Q84U55	Q84U55 rickettsia
41	98.5	12.6	154	16	Q89J15	Q89J15 bradyrhizob
42	98.5	12.6	453	5	Q9N6M8	Q9N6M8 drosophila
43	98.5	12.6	738	5	O02402	O02402 pintada fu
44	98.5	12.6	1286	2	Q841Y5	Q841Y5 campylobact
45	98.5	12.6	3659	16	Q98LN6	Q98LN6 rhizobium 1

#### ALIGNMENTS

RESULT 1

O33802 ID O33802 PRELIMINARY; PRT; 152 AA.

AC O33802; 01-JAN-1998 (TrEMBLrel. 05, Created)

DT 01-JAN-1998 (TrEMBLrel. 05, Last sequence update)

DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)

DE AGFA protein (Fragment).

GN AGFA

OS Salmonella typhimurium.

OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;

OC Enterobacteriaceae; Salmonella.

OX NCBI\_TaxID=602;

RN [1]

RP SEQUENCE FROM N.A.

RX MEDLINE=98053981; PubMed=9393832;

RA Sukupolvi S.S., Lorentz R.G., Gordon J.I., Bian Z., Pfeifer J.D.,

RA Normark S.J., Rhen M.;

RT "Expression of thin, aggregative fimbriae promotes interaction of

RT Salmonella typhimurium SR-11 with mouse small intestinal epithelial

RT cells.";

RL Infect. Immun. 65:5320-5325 (1997).

DR EMBL: AJ000514; CAA04151.1; -.

PT NON TER 152 152

SQ SEQUENCE 152 AA; 15401 MW; 9DA7DADC2364B006 CRC64;

Query Match 87.3%; Score 683; DB 2; Length 152;

Best Local Similarity 89.4%; Pred. No. 2.8e-48;

Matches 135; Conservative 3; Mismatches 13; Indels 0; Gaps 0;

Qy 1 MKLLKVAFAAIVVSGSALAGVVPQWGGGNHNGGNSGPDSTLSIYQYGSNAALYDQ 60

Db 1 MKLLKVAFAAIVVSGSALAGVVPQWGGGNHNGGNSGPDSTLSIYQYGSNAALALQ 60

Qy 61 LVTRVWTHVTHMAHAGYNGGADVGGQADNSTIELTQNGFRNNATIDQWNAKNSDITVGQYG 120

Db 61 SDARKSETTITQSGYNGADVGGQADNSTIELTQNGFRNNATIDQWNAKNSDITVGQYG 120

Qy 121 NNAALVNQTSADSSVMVRQVGFNNATANQY 151



RT "Production of Cellulose and Curli Fimbriae by Members of the Family  
RT Enterobacteriaceae Isolated from the Human Gastrointestinal Tract."  
RL Infect. Immun. 72:4151-4158 (2003).  
DR EMBL; AJ515702; CAD56678.1; -.  
SQ SEQUENCE 150 AA; 15112 MW; 5D8BB2D872DF15F3 CRC64;

Query Match 56.6%; Score 442.5; DB 2; Length 150;  
Best Local Similarity 60.3%; Pred. No. 1.2e-28;  
Matches 91; Conservative 25; Mismatches 34; Indels 1; Gaps 1;

QY 1 MKLLKVAAPAAIVVSGSALAGVVPQWGGGNGGNSGPDSTLSIYQYGSANAALYDQ 60  
Db 1 MKFIKVAALAAIVVSGSAMAGMNG-CGWGHGHGGYGGFNPSTLNIYQNGGNSALALQ 59  
QY 61 LVTRVVTHEMAHAGYNGADYCGGADNSTIELTQNGFNATIDQWNAKNSDITVGYGG 120  
Db 60 TDARNSVLNISTQGGNGADYCGGSDSSINLTQNGFNGSATLDQWNSKDSVMWVSQYGG 119  
QY 121 NNAALVNQTASDSSVMVRQVGFNNATANQY 151  
Db 120 LINGALVDQTASNSTVNTVQIGFGNHATAHQY 150

RESULT 6  
Q54069  
ID Q54069 PRELIMINARY; PRT; 76 AA.  
AC Q54069;  
DT 01-NOV-1996 (TrEMBLrel. 01, Created)  
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)  
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)  
DE SEF17 fimbria (Fragment).  
GN AGFA.  
OS Salmonella enteritidis.  
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;  
OC Enterobacteriaceae; Salmonella.  
OX NCBI\_TaxID=592;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=SE30;  
RA Cox J.M., Egglez S., Woolcock J.B.;  
RT "Virulence of Salmonella enteritidis in chickens correlates with  
RT colony morphology and expression of SEF17 fimbriae."  
RL Submitted (APR-1996) to the EMBL/GenBank/DBSJ databases.  
DR EMBL; U53207; AAA98671.1; -.  
FT NON TER 1  
FT NON TER 76  
SQ SEQUENCE 76 AA; 7704 MW; 2FD5411241A7BCB1 CRC64;

Query Match 38.9%; Score 304; DB 2; Length 76;  
Best Local Similarity 80.3%; Pred. No. 1e-17; Indels 0; Gaps 0;  
Matches 61; Conservative 2; Mismatches 13; Indels 0; Gaps 0;

QY 30 GNNHGGGNSGPDSTLSIYQYGSANAALYDQLVTRVVTHEMAHAGYNGADYCGGADNST 89  
Db 1 GNNHGGGNSGPDSTLSIYQYGSANAALALQSDARKSETTITQSGYNGADYCGGADNST 60  
QY 90 IELTQNGFRNNATIDQ 105  
Db 61 IELTQNGFRNNATIDQ 76

RESULT 7  
Q983J5  
ID Q983J5 PRELIMINARY; PRT; 29 AA.  
AC Q983J5;  
DT 01-MAY-2000 (TrEMBLrel. 13, Created)  
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)  
DT 01-MAR-2001 (TrEMBLrel. 16, Last annotation update)  
DE Curlin subunit monomer (Fragment).  
GN CSGA.  
OS Escherichia coli.  
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;  
OC Enterobacteriaceae; Escherichia.

OX NCBI\_TaxID=562;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC TRANSPOSON=insertion sequence IS1;  
RX MEDLINE=99314153; PubMed=10386375;  
RA La Ragione R.M., Collighan R.J., Woodward M.J.;  
RT "Non-cultivation of Escherichia coli O78:k80 isolates associated with  
RT IS1 inserti on in csng and reduced persistence in poultry infection."  
RL FEMS Microbiol. Lett. 175:247-253 (1999).  
DR EMBL; AJ131756; CAB5380.1; -.  
FT NON TER 29  
SQ SEQUENCE 29 AA; 2789 MW; E290DFC07ABBB243 CRC64;

Query Match 15.6%; Score 122; DB 2; Length 29;  
Best Local Similarity 89.7%; Pred. No. 0.0022;  
Matches 26; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 MKLLKVAAPAAIVVSGSALAGVVPQWGGG 29  
Db 1 MKLLKVAATAAIVFGSALAGVVPQYGG 29

RESULT 8  
Q8EIH4  
ID Q8EIH4 PRELIMINARY; PRT; 502 AA.  
AC Q8EIH4;  
DT 01-MAR-2003 (TrEMBLrel. 23, Created)  
DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)  
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)  
DE Conserved hypothetical protein.  
GN SO0865.  
OS Shewanella oneidensis.  
OC Bacteria; Proteobacteria; Gammaproteobacteria; Alteromonadales;  
OC Alteromonadaceae; Shewanella.  
OX NCBI\_TaxID=70863;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=MR-1;  
RX MEDLINE=22297686; PubMed=12368813;  
RA Heidelberg J.F., Paulsen I.T., Nelson K.E., Gaidos E.J., Nelson W.C.,  
RA Read T.D., Eisen J.A., Seshadri R., Ward M., Methe B., Clayton R.A.,  
RA Meyer T., Tsapin A., Scott J., Beaman M., Brinkac L., Daugherty S.,  
RA DeBoy R.T., Dodson R.J., Durkin A.S., Haft D.H., Kolonay J.F.,  
RA Madupu R., Peterson J.D., Unayam L.A., White O., Wolf A.M.,  
RA Vamathevan J., Weidman J., Impraim M., Lee K., Berry K., Lee C.,  
RA Mueller J., Khouri H., Gill J., Utterback T.R., McDonald L.A.,  
RA Feldblyum T.V., Smith H.O., Venter J.C., Nealon K.H., Fraser C.M.;  
RT "Genome sequence of the dissimilatory metal ion-reducing bacterium  
RT Shewanella oneidensis."  
RL Nat. Biotechnol. 20:1118-1123 (2002).  
DR EMBL; AB015532; AAN53441.1; -.  
DR TIGR; SO0865; -.  
KW Hypothetical protein; Complete proteome.  
SQ SEQUENCE 502 AA; 52441 MW; D08CA23D6C46B62D CRC64;

Query Match 15.5%; Score 121; DB 16; Length 502;  
Best Local Similarity 26.1%; Pred. No. 0.086;  
Matches 40; Conservative 22; Mismatches 57; Indels 34; Gaps 6;

QY 29 GGNHNG-----GGN-----SSGPDSTLSIYQYGSANAALYDQLVTR 64  
Db 231 GDNHTGFVYALAGSENDISMEQSGSNNTAYLSMTGTGTDNTVDITQDGSN-TVGDSLIAD 289  
QY 65 VVTHE-----MAHAGYNGADYCGGADNSTIELTQNGFRNNATIDQWNAKNSDITVGYGG 120  
Db 290 IQGDDNDITIKQKGDNSGAEFQWVGDSNDVLLKQKQGDANFATGAYGTDN-DFDLSKGD 348  
QY 121 NNAALVNQTASDSSVMVRQVGFNG----NATAN 149  
Db 349 NNELVAFATGEDNSIEISQEGDANFAVVDATGN 381

RESULT 9

```

Q8UGN9 PRELIMINARY; PRT; 145 AA.
ID Q8UGN9;
AC Q8UGN9;
DT 01-JUN-2002 (TrEMBLrel. 21, Created)
DT 01-JUN-2002 (TrEMBLrel. 21, Last sequence update)
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE Hypothetical protein Atu4768.
GN Atu4768 OR AGF_L_228.
OS Agrobacterium tumefaciens (strain C58 / ATCC 33970).
OC Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;
OC Rhizobiaceae; Rhizobium/Agrobacterium group; Agrobacterium.
OX NCBI_TaxID=176299;
RN [1]_
RP SEQUENCE FROM N.A.
RX MEDLINE=21608550; PubMed=11743193;
RA Wood D.W., Setubal J.C., Kaul R., Monks D.E., Kitajima J.P.,
RA Okura V.K., Zhou Y., Chen L., Wood G.E., Almeida N.F. Jr., Woo L.,
RA Chen Y., Paulsen I.T., Eisen J.A., Karp P.D., Bovee D. Sr.,
RA Chapman P., Clendenning J., Deatherage G., Gillet W., Grant C.,
RA Kutayavin T., Levy R., Li M.-J., McLelland E., Palmieri A.,
RA Raymond C., Rouse G., Saenphimmachak C., Wu Z., Romero P., Gordon D.,
RA Zhang S., Yoo H., Tao Y., Biddle P., Jung M., Krespan W., Perry M.,
RA Gordon-Kamm B., Liao L., Kim S., Hendrick C., Zhao Z.-Y., Dolan M.,
RA Chumley F., Tingey S.V., Tomb J.-F., Gordon M.P., Olson M.V.,
RA Nester E.W.;
RT "The genome of the natural genetic engineer Agrobacterium tumefaciens
RT C58".
RL Science 294:2317-2323(2001).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=21608551; PubMed=11743194;
RA Goodner B., Hinkle G., Gattung S., Miller N., Blanchard M.,
RA Currelo B., Goldman B.S., Cao Y., Askenazi M., Halling C., Mullin L.,
RA Houniel K., Gordon J., Vaudin M., Iatchouk O., Epp A., Liu F.,
RA Wollam C., Allinger M., Doughty D., Scott C., Lappas C., Markelz B.,
RA Flanagan C., Crowell C., Gursion J., Lomo C., Sear C., Strub G.,
RA Cielo C., Slater S.;
RT "Genome sequence of the plant pathogen and biotechnology agent
RT Agrobacterium tumefaciens C58".
RL Science 294:2323-2328(2001).
DR EMBL; AE009405; AL45562.1; -
DR EMBL; AE008209; AAK8682.1; -
DR FIR; AD3143; AD3143.
DR PIR; H98144; H98144.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 145 AA; 14984 MW; DEDC870E1713D51A CRC64;

Query Match 14.6%; Score 114.5; DB 16; Length 145;
Best Local Similarity 25.2%; Pred. No. 0.064;
Matches 39; Conservative 26; Mismatches 59; Indels 31; Gaps 4;

QY 3 LKVAAPAAIVVGSALAGVVPWG-----GGNHGNGNSGPDSTLSIYQVGSANA 55
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1 MIRKSPIAGALVALVGLSAAAPAMANDVRIEYQVGSNSAGGAQEGYGNRIITYQNGYN- 59
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :

QY 56 ALYDQLVTAVTHMAHAGVNGADVGGADNSTIETQNGFRNATIDQWNAKNSDITV 115
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 60 -----RIVGHQY---GRHNLAVGQGHNYGSTQNGNRNAGI----- 96
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :

QY 116 GQYGGNNAALVNTASDSSVMVRQVFGNNATANQ 150
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 97 GQFGSNHTTILTDGNGNIAAGVQVGRGCSANVSQ 131
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :

RESULT 10
Q92U08 PRELIMINARY; PRT; 2174 AA.
ID Q92U08;
AC Q92U08;
DT 01-DEC-2001 (TrEMBLrel. 19, Created)
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Hypothetical glycine-rich protein SMB21548.
GN RB0989 OR SMB21548.

```

```

OS Rhizobium meliloti (Sinorhizobium meliloti).
OG Plasmid pSymb (megaplasmid 2).
OC Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;
OC Rhizobiaceae; Sinorhizobium/Ensifer group; Sinorhizobium.
OX NCBI_TaxID=382;
RN [1]_
RP SEQUENCE FROM N.A.
RX STRAIN=1021;
RX MEDLINE=21396508; PubMed=11481431;
RA Finan T.M., Weidner S., Wong K., Buhrmester J., Chain P.,
RA Vorhoeffer F.J., Hernandez-Lucas I., Becker A., Gouzy J.,
RA Golding B., Puhler A.;
RT "The complete sequence of the 1,683-kb pSymb megaplasmid from the N2-
RT fixing endosymbiont Sinorhizobium meliloti."
RL Proc. Natl. Acad. Sci. U.S.A. 98:9889-9894(2001).
DR EMBL; AL603645; CAC49389.1; -
DR PIR; E95965; E95965.
DR GO; GO:0046821; C:extrachromosomal DNA; IEA.
DR GO; GO:0004601; F:peroxidase activity; IEA.
DR GO; GO:0006979; P:response to oxidative stress; IEA.
DR InterPro; IPR005546; Autotransporter.
DR InterPro; IPR002016; Peroxidase.
DR InterPro; IPR002173; Pfkb.
DR Pfam; PF03757; Autotransporter; 1.
DR PROSITE; PS00435; PEROXIDASE_1; 1.
DR PROSITE; PS00583; PFKB_KINASE_1; 2.
DR Plasmid; Hypothetical protein; Complete proteome.
KW Plasmid;
SQ SEQUENCE 2174 AA; 203314 MW; 008EB68297B44182 CRC64;

Query Match 14.5%; Score 113; DB 16; Length 2174;
Best Local Similarity 27.0%; Pred. No. 2.3;
Matches 40; Conservative 21; Mismatches 51; Indels 36; Gaps 7;

QY 11 AIVVGSALAGVVPQ--WGGNGNHGNGNSGPDSTLSIYQVGSANAALYDQLVTRVTH 68
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 693 AATAGAGAVGILAQSIGGGGN---GGNATGADGFGSFGIQQGGGG----- 737
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :

QY 69 EMAHAGVNGADV-----QGADNSTI--ELTQNGFRNATIDQWNAK---NSDITV 115
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 738 -----GGYANTANVVFGLTLITQGHAGVAGVSGGGGTGTASSYAG:GFTASNAV 793
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :

QY 116 GQYGGNNA--LVNQTASDSSVMVRQVQ 141
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 794 GGTGGGGAGGGEVSVSLTDSAIRTGQGG 821
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :

RESULT 11
Q7UCZ1 PRELIMINARY; PRT; 151 AA.
ID Q7UCZ1;
AC Q7UCZ1;
DT 01-OCT-2003 (TrEMBLrel. 25, Created)
DT 01-OCT-2003 (TrEMBLrel. 25, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Minor curlin subunit.
GN CSGB OR S1108.
OS Shigella flexneri.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Shigella.
OX NCBI_TaxID=623;
RN [1]_
RP SEQUENCE FROM N.A.
RX STRAIN=2457T / ATCC 700930 / Serotype 2a;
RX MEDLINE=22590274; PubMed=12704152;
RA Wei J., Goldberg M.B., Burland V., Venkatesan M.M., Deng W.,
RA Fournier G., Mayhew G.F., Plunkett G. III, Rose D.J., Darling A.,
RA Mau B., Perna N.T., Payne S.M., Runyen-Janecky L.J., Zhou S.,
RA Schwartz D.C., Blattner F.R.;
RT "Complete genome sequence and comparative genomics of Shigella
RT flexneri serotype 2a strain 2457T."
RL Infect. Immun. 71:2775-2786(2003).
DR EMBL; AE016981; AAP16542.1; -
DR EMBL; AE016981; AAP16542.1; -
SQ SEQUENCE 151 AA; 15868 MW; 5D5D266B964014A0 CRC64;

```



```
Query Match 14.3%; Score 112; DB 16; Length 151;
Best Local Similarity 29.8%; Pred. No. 0.11;
Matches 34; Conservative 17; Mismatches 14; Gaps 3;

QY 51 GSANAALYDQVLT--RVVTHEMAHAGYNGADVGGADNSTIELTQNGFR----- 98
DB 17 GIAAAGYDLANSEYNAFVNLKSSFNQAIIIGQAGTNNSAQLRQGGSKLLVVAQEGS 76
QY 99 -NNATIDQWNAKNSDITVQYGGNNAALVNOTASDSSVMVRQVGFNNATANQY 151
DB 77 SNRAKIDQTGDYNL-AYIDQAGSANDASISQAGYNTAMIIQKSGNKANITQY 129

RESULT 12
Q8CW64 PRELIMINARY; PRT; 160 AA.
AC Q8CW64;
DT 01-MAR-2003 (TrEMBLrel. 23, Created)
DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DE 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE Minor curlin subunit precursor.
GN CSGB OR C1305.
OS Escherichia coli O6.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Escherichia.
OX NCBI_TaxID=217992;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=O6:H1 / CFT073 / ATCC 700928;
RX MEDLINE=22388234; PubMed=12471157;
RA Welch R.A., Burland V., Plunkett G. III, Redford P., Roesch P.,
RA Rasko D., Buckles E.L., Liou S.-R., Boutin A., Hackett J., Stroud D.,
RA Mayhew G.F., Rose D.J., Zhou S., Schwartz D.C., Perna N.T.,
RA Mobley H.L.T., Donnenberg M.S., Blattner F.R.;
RT "Extensive mosaic structure revealed by the complete genome sequence
RT of uropathogenic Escherichia coli.";
RL Proc. Natl. Acad. Sci. U.S.A. 99:17020-17024(2002).
DR EMBL; AE016759; AAN79778.1; -.
KW Complete proteome.
SQ SEQUENCE 160 AA; 16963 MW; 49F68448D979B986 CRC64;

Query Match 14.3%; Score 112; DB 16; Length 160;
Best Local Similarity 29.8%; Pred. No. 0.12;
Matches 34; Conservative 17; Mismatches 49; Indels 14; Gaps 3;

QY 51 GSANAALYDQVLT--RVVTHEMAHAGYNGADVGGADNSTIELTQNGFR----- 98
DB 26 GIAAAGYDLANSEYNAFVNLKSSFNQAIIIGQAGTNNSAQLRQGGSKLLTVAQEGS 85
QY 99 -NNATIDQWNAKNSDITVQYGGNNAALVNOTASDSSVMVRQVGFNNATANQY 151
DB 86 SNRAKIDQTGDYNL-AYIDQAGSANDASISQAGYNTAMIIQKSGNKANITQY 138

RESULT 13
Q83RU7 PRELIMINARY; PRT; 160 AA.
AC Q83RU7;
DT 01-JUN-2003 (TrEMBLrel. 24, Created)
DT 01-JUN-2003 (TrEMBLrel. 24, Last sequence update)
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE Minor curlin subunit precursor, similar to CsgA.
GN CSGB OR SF1035.
OS Shigella flexneri.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Shigella.
OX NCBI_TaxID=623;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=301 / Serotype 2a;
RX MEDLINE=2272406; PubMed=12384590;
RA Jin Q., Yuan Z., Xu J., Wang Y., Shen Y., Lu W., Wang J., Liu H.,
RA Yang J., Yang F., Zhang X., Zhang J., Yang G., Wu H., Qu D., Dong J.,
```

```
RA Sun L., Xue Y., Zhao A., Gao Y., Zhu J., Kan B., Ding K., Chen S.,
RA Cheng H., Yao Z., He B., Chen R., Ma D., Qiang B., Wen Y., Hou Y.,
RA Yu J.;
RT "Genome sequence of Shigella flexneri 2a: insights into pathogenicity
RL Nucleic Acids Res. 30:4432-4441(2002).
DR EMBL; AE015131; AAN42658.1; -.
KW Complete proteome.
SQ SEQUENCE 160 AA; 16919 MW; 50269F5268D2A32F CRC64;

Query Match 14.3%; Score 112; DB 16; Length 160;
Best Local Similarity 29.8%; Pred. No. 0.12;
Matches 34; Conservative 17; Mismatches 49; Indels 14; Gaps 3;

QY 51 GSANAALYDQVLT--RVVTHEMAHAGYNGADVGGADNSTIELTQNGFR----- 98
DB 26 GIAAAGYDLANSEYNAFVNLKSSFNQAIIIGQAGTNNSAQLRQGGSKLLVVAQEGS 85
QY 99 -NNATIDQWNAKNSDITVQYGGNNAALVNOTASDSSVMVRQVGFNNATANQY 151
DB 86 SNRAKIDQTGDYNL-AYIDQAGSANDASISQAGYNTAMIIQKSGNKANITQY 138

RESULT 14
Q7X244 PRELIMINARY; PRT; 151 AA.
AC Q7X244;
DT 01-OCT-2003 (TrEMBLrel. 25, Created)
DT 01-OCT-2003 (TrEMBLrel. 25, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Nucleation component of curlin monomers.
GN CSGB.
OS Citrobacter sp. Fec2.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Citrobacter.
OX NCBI_TaxID=213763;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Fec2;
RA Zogaj X., Bokranz W., Nimitz M., Romling U.;
RT "Production of Cellulose and Curli Fimbriae by Members of the Family
RT Enterobacteriaceae Isolated from the Human Gastrointestinal Tract.";
RL Infect. Immun. 72:4151-4158(2003).
DR EMBL; AJ515700; CAD56671.1; -.
SQ SEQUENCE 151 AA; 16158 MW; BD00AF57E1400704 CRC64;

Query Match 14.1%; Score 110; DB 2; Length 151;
Best Local Similarity 27.2%; Pred. No. 0.16;
Matches 31; Conservative 22; Mismatches 47; Indels 14; Gaps 3;

QY 51 GSANAALYDQVLT--RVVTHEMAHAGYNGADVGGADNSTIELTQNGFR----- 98
DB 17 GIASATSYDLAHSEYNAFVNLKSSFNQAIIIGQAGTNNSAKVRQEGSKLLSVSQEGG 76
QY 99 -NNATIDQWNAKNSDITVQYGGNNAALVNOTASDSSVMVRQVGFNNATANQY 151
DB 77 SNRAKVDQSGAYNF-AYIAQSGHSNDASISQSNYNTAMIIQKSGNKANITQY 129

RESULT 15
Q8PD38 PRELIMINARY; PRT; 1333 AA.
AC Q8PD38;
DT 01-OCT-2002 (TrEMBLrel. 22, Created)
DT 01-OCT-2002 (TrEMBLrel. 22, Last sequence update)
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE Ice nucleation protein.
GN XCC0507.
OS Xanthomonas campestris (pv. campestris).
OC Bacteria; Proteobacteria; Gammaproteobacteria; Xanthomonadales;
OC Xanthomonadaceae; Xanthomonas.
OX NCBI_TaxID=340;
RN [1]
```

Search completed: August 2, 2004, 14:54:41  
Job time : 30.7 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: August 2, 2004, 14:35:42 ; Search time 44.9 Seconds  
(without alignments)  
950.215 Million cell updates/sec

Title: US-09-543-407-28

Perfect score: 775

Sequence: 1 MKLLKVAFAAIVVGSALA.....DSSVMVRQVGGNNATANYQ 151

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1586107 seqs, 282547505 residues

Total number of hits satisfying chosen parameters: 1586107

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : A\_Geneseq\_29Jan04:\*\*

- 1: Geneseq1980s:\*\*
- 2: Geneseq1990s:\*\*
- 3: Geneseq2000s:\*\*
- 4: Geneseq2001s:\*\*
- 5: Geneseq2002s:\*\*
- 6: Geneseq2003as:\*\*
- 7: Geneseq2003bs:\*\*
- 8: Geneseq2004s:\*\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	775	100.0	151	3	AAB36354
2	712	91.9	151	3	AAB36351
3	683	88.1	151	2	AAR74625
4	683	88.1	151	3	AAB36341
5	678	87.5	151	2	AAM23570
6	662	85.4	151	3	AAB36350
7	617	79.6	151	3	AAB36353
8	613	79.1	151	3	AAB36355
9	605	78.1	151	3	AAB36346
10	603	77.8	151	3	AAB36347
11	601	77.5	151	3	AAB36349
12	600	77.4	151	3	AAB36352
13	568	73.3	151	3	AAB36348
14	509	65.7	151	3	AAB36343
15	504	65.0	151	7	ABR82651
16	498	64.3	120	2	AAR2761
17	498	64.3	120	2	AAR23569
18	435	56.1	142	2	AAR2664
19	359	46.3	122	2	AAR2663
20	147	19.0	45	3	AAB36316
21	132	17.0	22	3	AAB36318
22	113	14.6	24	7	ABR82644
23	111	14.3	22	3	AAB36322
24	111	14.3	22	3	AAB36327
25	111	14.3	22	3	AAB36337

26	109	14.1	23	3	AAB36340	Aab36340 Salmonell
27	109	14.1	23	3	AAB36324	Aab36324 Salmonell
28	109	14.1	23	3	AAB36319	Aab36319 Salmonell
29	102	13.2	26	7	ABR82649	AbR82649 E. coli V
30	96	12.4	19	3	AAB36323	Aab36323 Salmonell
31	96	12.4	19	3	AAB36336	Aab36336 Salmonell
32	96	12.4	19	3	AAB36328	Aab36328 Salmonell
33	94.5	12.2	151	3	AAB36342	Aab36342 Salmonell
34	92	11.9	1074	6	ABU22692	Abu22692 Protein e
35	91.5	11.8	502	2	AAM232312	Aaw32312 Leishmani
36	90	11.6	24	7	ABR82642	AbR82642 E. coli N
37	89.5	11.5	151	3	AAB36344	Aab36344 Escherich
38	89	11.5	24	7	ABR82647	AbR82647 E. coli C
39	89	11.5	354	7	ABO23520	Abo23520 Mycobacte
40	87.5	11.3	249	3	AAY69523	Aay69523 Anti-CD38
41	87.5	11.3	447	3	AAG29728	Aag29728 Arabidops
42	87.5	11.3	468	3	AAG29727	Aag29727 Arabidops
43	87.5	11.3	842	5	ABP66189	Abp66189 Bifidobac
44	87.5	11.3	1419	5	ABP69842	Abp69842 Human pol
45	87.5	11.3	1477	5	ABP69841	Abp69841 Human pol

ALIGNMENTS

RESULT 1  
AAB36354  
ID AAB36354 standard; protein; 151 AA.  
XX  
AC AAB36354;  
XX  
DT 26-FEB-2001 (first entry)  
XX  
DE Agfa::PT3#9 amino acid sequence SEQ ID NO:28.  
XX  
KW Salmonella; agfa; chromosomal gene replacement; fimbrin; epitope;  
KW vaccine; immune response; immunogen.  
XX  
OS Salmonella enteritidis.  
OS Escherichia coli.  
OS Synthetic.  
XX  
PN WO2000060102-A2.  
XX  
PD 12-OCT-2000.  
XX  
PF 05-APR-2000; 2000WO-CA000356.  
XX  
PR 05-APR-1999; 99US-0127888P.  
XX  
(UYVI-) UNIV VICTORIA.  
XX  
White AP, Doran JL, Collison SK, Kay WW;  
WPI; 2000-672631/65.  
N-PSDB; AAC64630.  
Recombinant agfa gene having a segment replaced by a foreign DNA sequence  
which encodes foreign epitope or antigen, expresses recombinant Agfa  
protein useful for eliciting immune response in animal.  
Disclosure; Page 138; 139pp; English.

The present invention describes a recombinant agfa gene (I) where a segment of the gene has been replaced by a segment of a foreign DNA sequence which encodes a foreign epitope or antigen. Also described are: (1) use of thin aggregative fimbriae (SEF17/TAF) nucleation depended assembly system of strains of Salmonella, Escherichia coli and Enterobacteriaceae for the production of fimbriae comprising recombinant Agfa, CsgA and Agfa-homologue fimbriin subunits, respectively; (2) directing recombination of a recombinant gene into the chromosome of the homologous species; (3) directing recombination of a recombinant gene back into the chromosome of the homologous species, replacing the native

CC copy of that gene; and (4) eliciting an immune response in an animal,  
 CC comprising separating an amino acid polymer comprising a recombinant Agfa  
 CC protein containing a replacement segment or segments of foreign amino  
 CC acid sequence or sequences grown on a Salmonella, E. coli or  
 CC Enterobacteriaceae host cell, from the host cell and introducing the  
 CC polymer into the animal in conjunction with a carrier or diluent. (1) is  
 CC useful for the expression of recombinant Agfa protein which is useful for  
 CC eliciting an immune response in an animal. In a fimbrial presentation  
 CC system the heterologous antigens are presented in high numbers (up to  
 CC 500,000 copies/cell), the hybrid fimbrial protein possesses both the  
 CC immunogenicity and adhesion properties relevant for an efficient live  
 CC vaccine, the carrier fimbrial subunit proteins are usually strong  
 CC immunogens, which may be important for directing an immune response  
 CC against the inserted epitope, and hybrid fimbriae are easy and  
 CC inexpensive to purify in large amount. The present sequence is given in  
 CC the exemplification of the present invention  
 XX Sequence 151 AA;

Query Match 100.0%; Score 775; DB 3; Length 151;  
 Best Local Similarity 100.0%; Pred. No. 2.5e-67;  
 Matches 151; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MKLLKVAAPAAIVVSGSALAGVVPQWGGGNGGNSGPDSTLSIYQYGSANAALALQ 60  
 DB 1 MKLLKVAAPAAIVVSGSALAGVVPQWGGGNGGNSGPDSTLSIYQYGSANAALALQ 60

QY 61 SDARKSETTITQSGYNGADYDQLVTRVVTTHMAHAFRNATIDOWNAKNSDITVGOYGG 120  
 DB 61 SDARKSETTITQSGYNGADYDQLVTRVVTTHMAHAFRNATIDOWNAKNSDITVGOYGG 120

QY 121 NNAALVNQTASDSSVMVRQVFGNNATANQY 151  
 DB 121 NNAALVNQTASDSSVMVRQVFGNNATANQY 151

## RESULT 2

AAB36351  
 ID AAB36351 standard; protein; 151 AA.  
 AC AAB36351;  
 XX  
 XX 26-FEB-2001 (first entry)  
 DT  
 DE Agfa::PT3#6 amino acid sequence SEQ ID NO:22.  
 DE  
 KW Salmonella; agfa; chromosomal gene replacement; fimbria; epitope;  
 KW vaccine; immune response; immunogen.  
 XX  
 XX Salmonella enteritidis.  
 OS Escherichia coli.  
 OS Synthetic.  
 XX  
 XX WO200060102-A2.  
 XX  
 XX 12-OCT-2000.  
 XX  
 XX 05-APR-2000; 2000WO-CA000356.  
 XX  
 XX 05-APR-1999; 99US-0127888P.  
 XX  
 XX (UYVI-) UNIV VICTORIA.  
 XX  
 XX White AP, Doran JL, Collison SK, Kay WW;  
 XX  
 XX WPI; 2000-672631/55.  
 DR N-PSDB; AAC64627.  
 DR  
 XX  
 XX Recombinant agfa gene having a segment replaced by a foreign DNA sequence  
 PT which encodes foreign epitope or antigen, expresses recombinant Agfa  
 PT protein useful for eliciting immune response in animal.  
 XX  
 XX Disclosure; Page 137; 139pp; English.

XX The present invention describes a recombinant agfa gene (1) where a  
 CC segment of the gene has been replaced by a segment of a foreign DNA  
 CC sequence which encodes a foreign epitope or antigen. Also described are:  
 CC (1) use of thin aggregative fimbriae (SEF17/TAF) nucleation depended  
 CC assembly system of strains of Salmonella, Escherichia coli and  
 CC Enterobacteriaceae for the production of fimbriae comprising recombinant  
 CC Agfa, CsaA and Agfa-homologue fimbrial subunits, respectively. (2)  
 CC directing recombination of a recombinant gene into the chromosome of the  
 CC homologous species; (3) directing recombination of a recombinant gene  
 CC back into the chromosome of the homologous species, replacing the native  
 CC copy of that gene; and (4) eliciting an immune response in an animal,  
 CC comprising separating an amino acid polymer comprising a recombinant Agfa  
 CC protein containing a replacement segment or segments of foreign amino  
 CC acid sequence or sequences grown on a Salmonella, E. coli or  
 CC Enterobacteriaceae host cell, from the host cell and introducing the  
 CC polymer into the animal in conjunction with a carrier or diluent. (1) is  
 CC useful for the expression of recombinant Agfa protein which is useful for  
 CC eliciting an immune response in an animal. In a fimbrial presentation  
 CC system the heterologous antigens are presented in high numbers (up to  
 CC 500,000 copies/cell), the hybrid fimbrial protein possesses both the  
 CC immunogenicity and adhesion properties relevant for an efficient live  
 CC vaccine, the carrier fimbrial subunit proteins are usually strong  
 CC immunogens, which may be important for directing an immune response  
 CC against the inserted epitope, and hybrid fimbriae are easy and  
 CC inexpensive to purify in large amount. The present sequence is given in  
 CC the exemplification of the present invention

XX Sequence 151 AA;

Query Match 91.9%; Score 712; DB 3; Length 151;  
 Best Local Similarity 91.1%; Pred. No. 3.3e-61;  
 Matches 144; Conservative 0; Mismatches 0; Indels 14; Gaps 2;

QY 1 MKLLKVAAPAAIVVSGSALAGVVPQWGGGNGGNSGPDSTLSIYQYGSANAALALQ 60  
 DB 1 MKLLKVAAPAAIVVSGSALAGVVPQWGGGNGGNSGPDSTLSIYQYGSANAALALQ 60

QY 61 SDARKSETTITQSGYNGADYDQLVTRVVTTHMAHAFRNATIDOWNAKNSDI 113  
 DB 61 SDARKSETTITQSGYNGADYDQLVTRVVTTHMAHAFRNATIDOWNAKNSDI 113

QY 114 TVGQYGGNNAALVNQTASDSSVMVRQVFGNNATANQY 151  
 DB 114 TVGQYGGNNAALVNQTASDSSVMVRQVFGNNATANQY 151

## RESULT 3

AAR74625  
 ID AAR74625 standard; protein; 151 AA.  
 AC AAR74625;  
 XX  
 XX 25-MAR-2003 (revised)  
 DT 26-JUN-1995 (first entry)  
 DT  
 DE Agfa sequence.  
 DE  
 XX Salmonella; Agfa; vaccine.  
 KW Salmonella.  
 XX OS  
 XX WO9425598-A2.  
 XX  
 XX 10-NOV-1994.  
 XX  
 XX 26-APR-1994; 94WO-IB000207.  
 XX  
 XX 26-APR-1993; 93US-00054452.  
 XX  
 XX (UYVI-) UNIV VICTORIA INNOVATION & DEV CORP.  
 XX (KING/) KING J.  
 XX

PI Kay WW, Collinson SK, Clouthier SC, Doran JL;  
 XX WPI, 1994-358275/44.  
 DR N-PSDB; AAQ87467.  
 XX  
 PT Eliciting an immune response to Salmonella - using attenuated Salmonella  
 PT strains, vector constructs, or compens. contg. fimbrial type proteins.  
 XX  
 PS Disclosure; Fig 7B; 95pp; English.  
 XX  
 CC The Salmonella AgfA protein and DNA are used in vaccine and genetic  
 CC immunization compositions, respectively, to elicit an immune response to  
 CC Salmonella in animals (e.g. food producing animals) and humans. (Updated  
 CC on 25-MAR-2003 to correct PN field.)  
 XX  
 SQ Sequence 151 AA;  
 Query Match 88.1%; Score 683; DB 2; Length 151;  
 Best Local Similarity 90.1%; Pred. No. 2.2e-58;  
 Matches 136; Conservative 1; Mismatches 14; Indels 0; Gaps 0;  
 QY 1 MLLKVAFAAIVVSGSALAGVVPQWGGGNGHNGSSGPDSTLSIYQYGSANAALQ 60  
 DB 1 MLLKVAFAAIVVSGSALAGVVPQWGGGNGHNGSSGPDSTLSIYQYGSANAALQ 60  
 QY 61 SDARKSETTITQSGYNGADYDQLVTRVVTHEMAHFRNNATIDQWNAKNSDITVGYGG 120  
 DB 61 SDARKSETTITQSGYNGADVGQADNSTIETQNGFRNNATIDQWNAKNSDITVGYGG 120  
 QY 121 NNAALVNQTASDSSVMVRQVFGNNATANQY 151  
 DB 121 NNAALVNQTASDSSVMVRQVFGNNATANQY 151  
 RESULT 4  
 AAB36341  
 ID AAB36341 standard; protein; 151 AA.  
 AC AAB36341;  
 XX  
 DT 26-FEB-2001 (first entry)  
 DE Salmonella enteritidis AgfA amino acid sequence SEQ ID NO:5.  
 KW Salmonella; agfA; chromosomal gene replacement; fimbria; epitope;  
 KW vaccine; immune response; immunogen.  
 OS Salmonella enteritidis.  
 XX  
 PN WO200060102-A2.  
 XX  
 PD 12-OCT-2000.  
 XX  
 PF 05-APR-2000; 2000WO-CA000356.  
 XX  
 PR 05-APR-1999; 99US-0127889P.  
 XX  
 PA (UYVI-) UNIV VICTORIA.  
 XX  
 PI White AP, Doran JL, Collinson SK, Kay WW;  
 XX  
 DR WPI, 2000-672631/65.  
 DR N-PSDB; AAC64617.  
 XX  
 PT Recombinant agfA gene having a segment replaced by a foreign DNA sequence  
 PT which encodes foreign epitope or antigen, expresses recombinant AgfA  
 PT protein useful for eliciting immune response in animal.  
 XX  
 PS Disclosure; Page 135; 139pp; English.  
 XX  
 CC The present invention describes a recombinant agfA gene (I) where a  
 CC segment of the gene has been replaced by a segment of a foreign DNA  
 CC sequence which encodes a foreign epitope or antigen. Also described are:

CC (1) use of thin aggregative fimbriae (SEF17/TAF) nucleation depended  
 CC assembly system of strains of Salmonella, Escherichia coli and  
 CC Enterobacteriaceae for the production of fimbriae comprising recombinant  
 CC AgfA, CsgA and AgfA-homologue fimbria subunits, respectively; (2)  
 CC directing recombination of a recombinant gene into the chromosome of the  
 CC homologous species; (3) directing recombination of a recombinant gene  
 CC back into the chromosome of the homologous species, replacing the native  
 CC copy of that gene; and (4) eliciting an immune response in an animal,  
 CC comprising separating an amino acid polymer comprising a recombinant AgfA  
 CC protein containing a replacement segment or segments of foreign amino  
 CC acid sequence or sequences grown on a Salmonella, E. coli or  
 CC Enterobacteriaceae host cell, from the host cell and introducing the  
 CC polymer into the animal in conjunction with a carrier or diluent. (1) is  
 CC useful for the expression of recombinant AgfA protein which is useful for  
 CC eliciting an immune response in an animal. In a fimbrial presentation  
 CC system the heterologous antigens are presented in high numbers (up to  
 CC 500,000 copies/cell), the hybrid fimbria protein possesses both the  
 CC immunogenicity and adhesion properties relevant for an efficient live  
 CC vaccine, the carrier fimbrial subunit proteins are usually strong  
 CC immunogens, which may be important for directing an immune response  
 CC against the inserted epitope, and hybrid fimbriae are easy and  
 CC inexpensive to purify in large amount. The present sequence is given in  
 CC the exemplification of the present invention  
 XX  
 SQ Sequence 151 AA;  
 Query Match 88.1%; Score 683; DB 3; Length 151;  
 Best Local Similarity 90.1%; Pred. No. 2.2e-58;  
 Matches 136; Conservative 1; Mismatches 14; Indels 0; Gaps 0;  
 QY 1 MLLKVAFAAIVVSGSALAGVVPQWGGGNGHNGSSGPDSTLSIYQYGSANAALQ 60  
 DB 1 MLLKVAFAAIVVSGSALAGVVPQWGGGNGHNGSSGPDSTLSIYQYGSANAALQ 60  
 QY 61 SDARKSETTITQSGYNGADYDQLVTRVVTHEMAHFRNNATIDQWNAKNSDITVGYGG 120  
 DB 61 SDARKSETTITQSGYNGADVGQADNSTIETQNGFRNNATIDQWNAKNSDITVGYGG 120  
 QY 121 NNAALVNQTASDSSVMVRQVFGNNATANQY 151  
 DB 121 NNAALVNQTASDSSVMVRQVFGNNATANQY 151  
 RESULT 5  
 AAW23570  
 ID AAW23570 standard; protein; 151 AA.  
 AC AAW23570;  
 XX  
 DT 25-MAR-2003 (revised)  
 DT 29-SEP-1997 (first entry)  
 DE Salmonella enteritidis 27655-3b agfA.  
 XX  
 KW Enteropathogenic bacteria; enterobacteria; S. enteritidis; antibody.  
 OS Salmonella enteritidis.  
 XX  
 PH Key Location/Qualifiers  
 FT Misc-difference 123  
 FT /note= "Encoded by GCC"  
 XX  
 PN USS635617-A.  
 XX  
 PD 03-JUN-1997.  
 XX  
 PF 26-APR-1994; 94US-00233788.  
 XX  
 PR 26-APR-1993; 93US-00054452.  
 XX  
 PA (UYVI-) UNIV VICTORIA INNOVATION & DEV CORP.  
 XX  
 PI Collinson SK, Kay WW, Doran JL;

```

XX WPI; 1997-309886/28.
XX N-PSDB; AAT74142.
XX
PT Isolated Salmonella gene agfa - used for diagnosis of Salmonella or
PT enteropathogenic bacteria of the Enterobacteria family.
XX
XX Example 2; Fig 7; 85pp; English.
XX
CC The present sequence represents agfa encoded by the full agfa gene
CC derived from Salmonella enteritidis 27655-3b. The nucleic acid can be
CC used to provide diagnostic assays for Salmonella and/or enteropathogenic
CC bacteria of the family Enterobacteria. It can also be used to provide
CC proteins and antibodies which can be used for assays. The nucleic acid
CC sequence can be used to provide probes or primers which can specifically
CC hybridise to nucleic acid molecules from greater than 99% of Salmonella
CC strains that are pathogenic to warm-blooded animals relative to nucleic
CC acid molecules from virtually all other microbial organisms. (Updated on
CC 23-MAR-2003 to correct PF field.)
XX
XX Sequence 151 AA;
SQ
    Query Match      87.5%; Score 678; DB 2; Length 151;
    Best Local Similarity 89.4%; Pred. No. 6.6e-58;
    Matches 135; Conservative 1; Mismatches 15; Indels 0; Gaps 0;
    QY 1 MLLKVAFAAIVVSGSALAGVVPQWGGGNGGNSGPDSTLSIYQYGSANAALALQ 60
    DB 1 MLLKVAFAAIVVSGSALAGVVPQWGGGNGGNSGPDSTLSIYQYGSANAALALQ 60
    QY 61 SDARKSETTITQSGYNGADYDQLVTRVVTHEMAHAFRNATIDQWNAKNSDITVGYGG 120
    DB 61 SDARKSETTITQSGYNGADYDQLVTRVVTHEMAHAFRNATIDQWNAKNSDITVGYGG 120
    QY 121 NNAALVNQTASDSSVMVRQVGFNNATANQY 151
    DB 121 NNPALVNQTASDSSVMVRQVGFNNATANQY 151
    RESULT 6
    ID AAB36350 standard; protein; 151 AA.
    AC AAB36350;
    DT 26-FEB-2001 (first entry)
    DE Agfa::PT3#5 amino acid sequence SEQ ID NO:20.
    KW Salmonella; agfa; chromosomal gene replacement; fimbrin; epitope;
    KW vaccine; immune response; immunogen.
    OS Salmonella enteritidis.
    OS Escherichia coli.
    OS Synthetic.
    QY WO200060102-A2.
    DB 12-OCT-2000.
    PF 05-APR-2000; 2000WO-CA000356.
    PR 05-APR-1999; 99US-0127888P.
    PA (UYVI-) UNIV VICTORIA.
    PI White AP, Doran JL, Collison SK, Kay WW;
    XX WPI; 2000-672631/65.
    DR N-PSDB; AAC64626.
    XX
    PT Recombinant agfa gene having a segment replaced by a foreign DNA sequence
    PT which encodes foreign epitope or antigen, expresses recombinant Agfa

```

```

PT protein useful for eliciting immune response in animal.
XX
XX Disclosure; Page 137; 139pp; English.
XX
CC The present invention describes a recombinant agfa gene (I) where a
CC segment of the gene has been replaced by a segment of a foreign DNA
CC sequence which encodes a foreign epitope or antigen. Also described are:
CC (1) use of thin aggregative fimbriae (SEF17/PAF) nucleation depended
CC assembly system of strains of Salmonella, Escherichia coli and
CC Enterobacteriaceae for the production of fimbriae comprising recombinant
CC Agfa, CsgA and Agfa-homologue fimbrial subunits, respectively; (2)
CC directing recombination of a recombinant gene into the chromosome of the
CC homologous species; (3) directing recombination of a recombinant gene
CC back into the chromosome of the homologous species, replacing the native
CC copy of that gene; and (4) eliciting an immune response in an animal
CC comprising separating an amino acid polymer comprising a recombinant Agfa
CC protein containing a replacement segment or segments of foreign amino
CC acid sequence or sequences grown on a Salmonella, E. coli or
CC Enterobacteriaceae host cell, from the host cell and introducing the
CC polymer into the animal in conjunction with a carrier or diluent. (I) is
CC useful for the expression of recombinant Agfa protein which is useful for
CC eliciting an immune response in an animal. In a fimbrial presentation
CC system the heterologous antigens are presented in high numbers (up to
CC 500,000 copies/cell), the hybrid fimbrial protein possesses both the
CC immunogenicity and adhesion properties relevant for an efficient live
CC vaccine, the carrier fimbrial subunit proteins are usually strong
CC immunogens, which may be important for directing an immune response
CC against the inserted epitope, and hybrid fimbriae are easy and
CC inexpensive to purify in large amount. The present sequence is given in
CC the exemplification of the present invention
XX
XX Sequence 151 AA;
SQ
    Query Match      85.4%; Score 662; DB 3; Length 151;
    Best Local Similarity 81.9%; Pred. No. 2.4e-56;
    Matches 136; Conservative 0; Mismatches 0; Indels 30; Gaps 2;
    QY 1 MLLKVAFAAIVVSGSALAGVVPQWGGGNGGNSGPDSTLSIYQYGSANAALALQ 60
    DB 1 MLLKVAFAAIVVSGSALAGVVPQWGGGNGGNSGPDSTLSIYQYGSANAALALQ 60
    QY 61 SDARKSETTITQSGYNGADYDQLVTRVVTHEMAHA-----FRNNATIDQ 105
    DB 61 SDARK-----YDQLVTRVVTHEMAHAGQAGDNSTIELTQNGFRNATIDQ 105
    QY 106 WNAKNSDITVGYGGNNAALVNQTASDSSVMVRQVGFNNATANQY 151
    DB 106 WNAKNSDITVGYGGNNAALVNQTASDSSVMVRQVGFNNATANQY 151
    RESULT 7
    ID AAB36353 standard; protein; 151 AA.
    AC AAB36353;
    DT 26-FEB-2001 (first entry)
    DE Agfa::PT3#8 amino acid sequence SEQ ID NO:26.
    KW Salmonella; agfa; chromosomal gene replacement; fimbrin; epitope;
    KW vaccine; immune response; immunogen.
    OS Salmonella enteritidis.
    OS Escherichia coli.
    OS Synthetic.
    QY WO200060102-A2.
    DB 12-OCT-2000.
    PF 05-APR-2000; 2000WO-CA000356.
    PT

```

PR 05-APR-1999; 99US-0127888P.  
 XX (UYVI-) UNIV VICTORIA.  
 PA White AP, Doran JL, Collison SK, Kay WW;  
 XX WPI; 2000-672631/65.  
 XX N-PSDB; AAC64629.  
 XX Recombinant agfa gene having a segment replaced by a foreign DNA sequence  
 PT which encodes foreign epitope or antigen, expresses recombinant Agfa  
 PT protein useful for eliciting immune response in animal.  
 XX Disclosure; Page 139; 139pp; English.  
 XX The present invention describes a recombinant agfa gene (1) where a  
 CC segment of the gene has been replaced by a segment of a foreign DNA  
 CC sequence which encodes a foreign epitope or antigen. Also described are:  
 CC (1) use of thin aggregative fimbriae (SEF17/TAFF) nucleation depended  
 CC assembly system of strains of Salmonella, Escherichia coli and  
 CC Enterobacteriaceae for the production of fimbriae comprising recombinant  
 CC Agfa, CsgA and Agfa-homologue fimbrial subunits, respectively; (2)  
 CC directing recombination of a recombinant gene into the chromosome of the  
 CC homologous species; (3) directing recombination of a recombinant gene  
 CC back into the chromosome of the homologous species, replacing the native  
 CC copy of that gene; and (4) eliciting an immune response in an animal,  
 CC comprising separating an amino acid polymer comprising a recombinant Agfa  
 CC protein containing a replacement segment or segments of foreign amino  
 CC acid sequence or sequences grown on a Salmonella, E. coli or  
 CC Enterobacteriaceae host cell, from the host cell and introducing the  
 CC polymer into the animal in conjunction with a carrier or diluent. (1) is  
 CC useful for the expression of recombinant Agfa protein which is useful for  
 CC eliciting an immune response in an animal. In a fimbrial presentation  
 CC system the heterologous antigens are presented in high numbers (up to  
 CC 500,000 copies/cell), the hybrid fimbrial protein possesses both the  
 CC immunogenicity and adhesion properties relevant for an efficient live  
 CC vaccine, the carrier fimbrial subunit proteins are usually strong  
 CC immunogens, which may be important for directing an immune response  
 CC against the inserted epitope, and hybrid fimbriae are easy and  
 CC inexpensive to purify in large amount. The present sequence is given in  
 CC the exemplification of the present invention  
 XX Sequence 151 AA;  
 SQ  
 Query Match 79.6%; Score 617; DB 3; Length 151;  
 Best Local Similarity 73.6%; Pred. No. 5.7e-52;  
 Matches 128; Conservative 0; Mismatches 0; Indels 46; Gaps 2;  
 QY 1 MLLKVAAPFAIIVVSGSALAGVVPQGGGNGHNGGNSGPDSTLSIYQYGSANAALALQ 60  
 DB 1 MLLKVAAPFAIIVVSGSALAGVVPQGGGNGHNGGNSGPDSTLSIYQYGSANAALALQ 57  
 QY 61 SDARKSETTITQSGYNGADYDQLVTRVVTHEMAHA-----F 97  
 DB 58 -----YDQLVTRVVTHEMAHAGYNGADVGQGDADNNTIELTQNGF 97  
 QY 98 RNNATIDQWAKNSDITVGYGGNNAALVNOTASDSSVMVQYGFQGNNTATQY 151  
 DB 98 RNNATIDQWAKNSDITVGYGGNNAALVNOTASDSSVMVQYGFQGNNTATQY 151  
 RESULT 8  
 AAB36355  
 ID AAB36355 standard; protein; 151 AA.  
 XX AAB36355;  
 AC AAB36355;  
 XX 26-FEB-2001 (first entry)  
 DT Agfa::PT3#10 amino acid sequence SEQ ID NO:30.  
 XX Salmonella; agfa; chromosomal gene replacement; fimbrial; epitope;  
 XX vaccine; immune response; immunogen.

XX Salmonella enteritidis.  
 OS Escherichia coli.  
 OS Synthetic.  
 XX WO2000060102-A2.  
 XX 12-OCT-2000.  
 XX 05-APR-2000; 2000WO-CA000356.  
 XX 05-APR-1999; 99US-0127888P.  
 XX (UYVI-) UNIV VICTORIA.  
 XX White AP, Doran JL, Collison SK, Kay WW;  
 XX WPI; 2000-672631/65.  
 XX N-PSDB; AAC64631.  
 XX Recombinant agfa gene having a segment replaced by a foreign DNA sequence  
 PT which encodes foreign epitope or antigen, expresses recombinant Agfa  
 PT protein useful for eliciting immune response in animal.  
 XX Disclosure; Page 139; 139pp; English.  
 XX The present invention describes a recombinant agfa gene (1) where a  
 CC segment of the gene has been replaced by a segment of a foreign DNA  
 CC sequence which encodes a foreign epitope or antigen. Also described are:  
 CC (1) use of thin aggregative fimbriae (SEF17/TAFF) nucleation depended  
 CC assembly system of strains of Salmonella, Escherichia coli and  
 CC Enterobacteriaceae for the production of fimbriae comprising recombinant  
 CC Agfa, CsgA and Agfa-homologue fimbrial subunits, respectively; (2)  
 CC directing recombination of a recombinant gene into the chromosome of the  
 CC homologous species; (3) directing recombination of a recombinant gene  
 CC back into the chromosome of the homologous species, replacing the native  
 CC copy of that gene; and (4) eliciting an immune response in an animal,  
 CC comprising separating an amino acid polymer comprising a recombinant Agfa  
 CC protein containing a replacement segment or segments of foreign amino  
 CC acid sequence or sequences grown on a Salmonella, E. coli or  
 CC Enterobacteriaceae host cell, from the host cell and introducing the  
 CC polymer into the animal in conjunction with a carrier or diluent. (1) is  
 CC useful for the expression of recombinant Agfa protein which is useful for  
 CC eliciting an immune response in an animal. In a fimbrial presentation  
 CC system the heterologous antigens are presented in high numbers (up to  
 CC 500,000 copies/cell), the hybrid fimbrial protein possesses both the  
 CC immunogenicity and adhesion properties relevant for an efficient live  
 CC vaccine, the carrier fimbrial subunit proteins are usually strong  
 CC immunogens, which may be important for directing an immune response  
 CC against the inserted epitope, and hybrid fimbriae are easy and  
 CC inexpensive to purify in large amount. The present sequence is given in  
 CC the exemplification of the present invention  
 XX Sequence 151 AA;  
 SQ  
 Query Match 79.1%; Score 613; DB 3; Length 151;  
 Best Local Similarity 74.6%; Pred. No. 1.4e-51;  
 Matches 129; Conservative 0; Mismatches 0; Indels 44; Gaps 2;  
 QY 1 MLLKVAAPFAIIVVSGSALAGVVPQGGGNGHNGGNSGPDSTLSIYQYGSANAALALQ 60  
 DB 1 MLLKVAAPFAIIVVSGSALAGVVPQGGGNGHNGGNSGPDSTLSIYQYGSANAALALQ 60  
 QY 61 SDARKSETTITQSGYNGAD-----YDQLVTRVVTHEMAHAF 98  
 DB 61 SDARKSETTITQSGYNGADVGQGDADNNTIELTQNGFRNNATYDQLVTRVVTHEMAHA-- 118  
 QY 99 RNNATIDQWAKNSDITVGYGGNNAALVNOTASDSSVMVQYGFQGNNTATQY 151  
 DB 119 -----GNNALVNOTASDSSVMVQYGFQGNNTATQY 151  
 RESULT 9

Qy Db



CC the exemplification of the present invention  
 XX Sequence 151 AA;  
 SQ

Query Match 77.8%; Score 603; DB 3; Length 151;  
 Best Local Similarity 80.8%; Pred. No. 1.3e-50;  
 Matches 122; Conservative 4; Mismatches 25; Indels 0; Gaps 0;

QY 1 MLLKLVAAFAAIVVSGSALAGVVPQWGGGNGHNGGNSGSPDSTLSIYQYGSANAALALQ 60  
 DB 1 MLLKLVAAFAAIVVSGSALAGVVPQWGGGNGHNGGNSGSPDSTLSIYQYGSANAALALQ 60

QY 61 SDARKSETTITQSGYNGADYDQLVTRVVTHEMAHAFRNATIDQWNAKNSDITVGYGG 120  
 DB 61 SDARKSETTITQSGYNGADVGGQADNSTIELTQNGFENNATIDQWNAKNSDITVGYDQ 120

QY 121 NNAALVNQTASDSVMVRQVGFNNATANQY 151  
 DB 121 LVTRVVTHEMAHASVMVRQVGFNNATANQY 151

RESULT 11  
 AAB36349  
 ID AAB36349 standard; protein; 151 AA.  
 XX  
 AC AAB36349;  
 XX  
 DT 26-FEB-2001 (first entry)  
 XX  
 DE Agfa::PT3#4 amino acid sequence SEQ ID NO:18.  
 XX  
 KW Salmonella; agfa; chromosomal gene replacement; fimbria; epitope;  
 KW vaccine; immune response; immunogen.  
 XX  
 OS Salmonella enteritidis.  
 OS Escherichia coli.  
 OS Synthetic.  
 XX  
 FN WO200060102-A2.  
 XX  
 PD 12-OCT-2000.  
 XX  
 PF 05-APR-2000; 2000WO-CA000356.  
 XX  
 PR 05-APR-1999; 99US-0127888P.  
 XX  
 PA (UUVI-) UNIV VICTORIA.  
 XX  
 PI White AP, Doran JL, Collison SK, Kay WW;  
 XX  
 DR WPI; 2000-672631/65.  
 XX  
 PT Recombinant agfa gene having a segment replaced by a foreign DNA sequence  
 PT which encodes foreign epitope or antigen, expresses recombinant Agfa  
 PT protein useful for eliciting immune response in animal.  
 XX  
 PS Disclosure; Page 136; 139pp; English.

The present invention describes a recombinant agfa gene (I) where a segment of the gene has been replaced by a segment of a foreign DNA sequence which encodes a foreign epitope or antigen. Also described are: (1) use of thin aggregative fimbriae (SEF17/TAF) nucleation depended assembly system of strains of Salmonella, Escherichia coli and Enterobacteriaceae for the production of fimbriae comprising recombinant Agfa, CsgA and Agfa-homologue fimbria subunits, respectively; (2) directing recombination of a recombinant gene into the chromosome of the homologous species; (3) directing recombination of a recombinant gene back into the chromosome of the homologous species, replacing the native copy of that gene; and (4) eliciting an immune response in an animal, comprising separating an amino acid polymer comprising a recombinant Agfa protein containing a replacement segment or segments of foreign amino acid sequence or sequences grown on a Salmonella, E. coli or

CC Enterobacteriaceae host cell, from the host cell and introducing the polymer into the animal in conjunction with a carrier or diluent. (I) is useful for the expression of recombinant Agfa protein which is useful for eliciting an immune response in an animal. In a fimbrial presentation system the heterologous antigens are presented in high numbers (up to 500,000 copies/cell), the hybrid fimbria protein possesses both the immunogenicity and adhesion properties relevant for an efficient live vaccine, the carrier fimbrial subunit proteins are usually strong immunogens, which may be important for directing an immune response against the inserted epitope, and hybrid fimbriae are easy and inexpensive to purify in large amount. The present sequence is given in the exemplification of the present invention

XX Sequence 151 AA;  
 SQ

Query Match 77.5%; Score 601; DB 3; Length 151;  
 Best Local Similarity 80.8%; Pred. No. 2e-50;  
 Matches 122; Conservative 4; Mismatches 25; Indels 0; Gaps 0;

QY 1 MLLKLVAAFAAIVVSGSALAGVVPQWGGGNGHNGGNSGSPDSTLSIYQYGSANAALALQ 60  
 DB 1 MLLKLVAAFAAIVVSGSALAGVVPQWGGGNGHNGGNSGSPDSTLSIYQYGSANAALALQ 60

QY 61 SDARKSETTITQSGYNGADYDQLVTRVVTHEMAHAFRNATIDQWNAKNSDITVGYGG 120  
 DB 61 SDARKSETTITQSGYNGADVGGQADNSTIELTQNGFENNATIDQWNAKNSDITVGYGG 120

QY 121 NNAALVNQTASDSVMVRQVGFNNATANQY 151  
 DB 121 NNAALVNQTASDSVMVRQVGFNNATANQY 151

RESULT 12  
 AAB36352  
 ID AAB36352 standard; protein; 151 AA.  
 XX  
 AC AAB36352;  
 XX  
 DT 26-FEB-2001 (first entry)  
 XX  
 DE Agfa::PT3#7 amino acid sequence SEQ ID NO:24.  
 XX  
 KW Salmonella; agfa; chromosomal gene replacement; fimbria; epitope;  
 KW vaccine; immune response; immunogen.  
 XX  
 OS Salmonella enteritidis.  
 OS Escherichia coli.  
 OS Synthetic.  
 XX  
 FN WO200060102-A2.  
 XX  
 PD 12-OCT-2000.  
 XX  
 PF 05-APR-2000; 2000WO-CA000356.  
 XX  
 PR 05-APR-1999; 99US-0127888P.  
 XX  
 PA (UUVI-) UNIV VICTORIA.  
 XX  
 PI White AP, Doran JL, Collison SK, Kay WW;  
 XX  
 DR WPI; 2000-672631/65.  
 XX  
 PT Recombinant agfa gene having a segment replaced by a foreign DNA sequence  
 PT which encodes foreign epitope or antigen, expresses recombinant Agfa  
 PT protein useful for eliciting immune response in animal.  
 XX  
 PS Disclosure; Page 138; 139pp; English.

The present invention describes a recombinant agfa gene (I) where a segment of the gene has been replaced by a segment of a foreign DNA sequence which encodes a foreign epitope or antigen. Also described are: (1) use of thin aggregative fimbriae (SEF17/TAF) nucleation depended assembly system of strains of Salmonella, Escherichia coli and Enterobacteriaceae for the production of fimbriae comprising recombinant Agfa, CsgA and Agfa-homologue fimbria subunits, respectively; (2) directing recombination of a recombinant gene into the chromosome of the homologous species; (3) directing recombination of a recombinant gene back into the chromosome of the homologous species, replacing the native copy of that gene; and (4) eliciting an immune response in an animal, comprising separating an amino acid polymer comprising a recombinant Agfa protein containing a replacement segment or segments of foreign amino acid sequence or sequences grown on a Salmonella, E. coli or

CC (1) use of thin aggregative fimbriae (SEF17/TAF) nucleation depended  
 CC assembly system of strains of Salmonella, Escherichia coli and  
 CC Enterobacteriaceae for the production of fimbriae comprising recombinant  
 CC AgfA, CsgA and AgfA-homologue fimbrial subunits, respectively; (2)  
 CC directing recombination of a recombinant gene into the chromosome of the  
 CC homologous species; (3) directing recombination of a recombinant gene  
 CC back into the chromosome of the homologous species, replacing the native  
 CC copy of that gene; and (4) eliciting an immune response in an animal,  
 CC comprising separating an amino acid polymer comprising a recombinant AgfA  
 CC protein containing a replacement segment or segments of foreign amino  
 CC acid sequence or sequences grown on a Salmonella, E. coli or  
 CC Enterobacteriaceae host cell, from the host cell and introducing the  
 CC polymer into the animal in conjunction with a carrier or diluent. (1) is  
 CC useful for the expression of recombinant AgfA protein which is useful for  
 CC eliciting an immune response in an animal. In a fimbrial presentation  
 CC system the heterologous antigens are presented in high numbers (up to  
 CC 500,000 copies/cell), the hybrid fimbrial protein possesses both the  
 CC immunogenicity and adhesion properties relevant for an efficient live  
 CC vaccine, the carrier fimbrial subunit proteins are usually strong  
 CC immunogens, which may be important for directing an immune response  
 CC against the inserted epitope, and hybrid fimbriae are easy and  
 CC inexpensive to purify in large amount. The present sequence is given in  
 CC the exemplification of the present invention

XX Sequence 151 AA;

Query Match 77.4%; Score 600; DB 3; Length 151;  
 Best Local Similarity 81.5%; Pred. No. 2.5e-50;  
 Matches 123; Conservative 3; Mismatches 25; Indels 0; Gaps 0;  
 QY 1 MKLLKVAAPAAIIVVSGSALAGVVPQWGGGNGGNGSSGPDSTLSIYQYGSANAALALQ 60  
 Db 1 MKLLKVAAPAAIIVVSGSALAGVVPQWGGGNGGNGSSGPDSTLSIYQYGSANAALALQ 60  
 QY 61 SDARKSETTITQSGYNGADYDQLVTRVVTHEMAHAFRNATIDQNAKNSDITVQYGG 120  
 Db 61 SDARKSETTITQSGYNGADYDQLVTRVVTHEMAHAFRNATIDQNAKNSDITVQYGG 120  
 QY 121 NNAALVNOTASDSSVMVRQVGFNNATANQY 151  
 Db 121 HEMAHANQATASDSSVMVRQVGFNNATANQY 151

RESULT 13  
 AAB36348  
 ID AAB36348 standard; protein; 151 AA.  
 AC AAB36348;  
 XX

DT 26-FEB-2001 (first entry)

XX AgfA::PT3#3 amino acid sequence SEQ ID NO:16.

DE Salmonella; agfA; chromosomal gene replacement; fimbria; epitope;  
 KW vaccine; immune response; immunogen.

XX Salmonella enteritidis.  
 OS Escherichia coli.  
 OS Synthetic.

XX WO200060102-A2.

XX 12-OCT-2000.

XX 05-APR-2000; 2000WO-CA000356.

XX 05-APR-1999; 99US-0127888P.

XX (UYVI-) UNIV VICTORIA.

XX White AP, Doran JL, Collison SK, Kay WW;

XX WPI; 2000-672631/65.

DR N-PSDB; AAC64624.

XX Recombinant agfA gene having a segment replaced by a foreign DNA sequence  
 PT which encodes foreign epitope or antigen, expresses recombinant AgfA  
 PT protein useful for eliciting immune response in animal.  
 XX Disclosure; Page 136; 139pp; English.

XX The present invention describes a recombinant agfA gene (I) where a  
 CC segment of the gene has been replaced by a segment of a foreign DNA  
 CC sequence which encodes a foreign epitope or antigen. Also described are:  
 CC (1) use of thin aggregative fimbriae (SEF17/TAF) nucleation depended  
 CC assembly system of strains of Salmonella, Escherichia coli and  
 CC Enterobacteriaceae for the production of fimbriae comprising recombinant  
 CC AgfA, CsgA and AgfA-homologue fimbrial subunits, respectively; (2)  
 CC directing recombination of a recombinant gene into the chromosome of the  
 CC homologous species; (3) directing recombination of a recombinant gene  
 CC back into the chromosome of the homologous species, replacing the native  
 CC copy of that gene; and (4) eliciting an immune response in an animal,  
 CC comprising separating an amino acid polymer comprising a recombinant AgfA  
 CC protein containing a replacement segment or segments of foreign amino  
 CC acid sequence or sequences grown on a Salmonella, E. coli or  
 CC Enterobacteriaceae host cell, from the host cell and introducing the  
 CC polymer into the animal in conjunction with a carrier or diluent. (I) is  
 CC useful for the expression of recombinant AgfA protein which is useful for  
 CC eliciting an immune response in an animal. In a fimbrial presentation  
 CC system the heterologous antigens are presented in high numbers (up to  
 CC 500,000 copies/cell), the hybrid fimbrial protein possesses both the  
 CC immunogenicity and adhesion properties relevant for an efficient live  
 CC vaccine, the carrier fimbrial subunit proteins are usually strong  
 CC immunogens, which may be important for directing an immune response  
 CC against the inserted epitope, and hybrid fimbriae are easy and  
 CC inexpensive to purify in large amount. The present sequence is given in  
 CC the exemplification of the present invention

XX Sequence 151 AA;

Query Match 73.3%; Score 568; DB 3; Length 151;  
 Best Local Similarity 80.1%; Pred. No. 3.3e-47;  
 Matches 121; Conservative 4; Mismatches 26; Indels 0; Gaps 0;

QY 1 MKLLKVAAPAAIIVVSGSALAGVVPQWGGGNGGNGSSGPDSTLSIYQYGSANAALALQ 60  
 Db 1 MKLLKVAAPAAIIVVSGSALAGVVPQWGGGNGGNGSSGPDSTLSIYQYGSANAALALQ 60  
 QY 61 SDARKSETTITQSGYNGADYDQLVTRVVTHEMAHAFRNATIDQNAKNSDITVQYGG 120  
 Db 61 SDARKSETTITQSGYNGADYDQLVTRVVTHEMAHAFRNATIDQNAKNSDITVQYGG 120  
 QY 121 NNAALVNOTASDSSVMVRQVGFNNATANQY 151  
 Db 121 NNAALVNOTASDSSVMVRQVGFNNATANQY 151

RESULT 14

AAB36343  
 ID AAB36343 standard; protein; 151 AA.  
 AC AAB36343;  
 XX

DT 26-FEB-2001 (first entry)

XX Escherichia coli CsgA amino acid sequence SEQ ID NO:7.  
 DE Salmonella; agfA; chromosomal gene replacement; fimbria; epitope;  
 KW vaccine; immune response; immunogen.

XX Escherichia coli.

XX WO200060102-A2.

XX 12-OCT-2000.

PF 05-APR-2000; 2000WO-CA000356.  
 PR 05-APR-1999; 99US-0127888P.  
 XX (UYVI-) UNIV VICTORIA.

PI White AP; Doran JL, Collison SK, Kay WW;  
 XX

XX WPI; 2000-672631/65.  
 DR N-PSDB; AAC64619.

XX Recombinant agfa gene having a segment replaced by a foreign DNA sequence  
 PT which encodes foreign epitope or antigen, expresses recombinant Agfa  
 PT protein useful for eliciting immune response in animal.

XX Disclosure; Page 135; 139pp; English.

XX The present invention describes a recombinant agfa gene (1) where a  
 CC segment of the gene has been replaced by a segment of a foreign DNA  
 CC sequence which encodes a foreign epitope or antigen. Also described are:  
 CC (1) use of thin aggregative fimbriae (SEF17/7AF) nucleation depended  
 CC assembly system of strains of Salmonella, Escherichia coli and  
 CC Enterobacteriaceae for the production of fimbriae comprising recombinant  
 CC Agfa, CsgA and Agfa-homologue fimbria subunits, respectively; (2)  
 CC directing recombination of a recombinant gene into the chromosome of the  
 CC homologous species; (3) directing recombination of a recombinant gene  
 CC back into the chromosome of the homologous species, replacing the native  
 CC copy of that gene; and (4) eliciting an immune response in an animal,  
 CC comprising separating an amino acid polymer comprising a recombinant Agfa  
 CC protein containing a replacement segment or segments of foreign amino  
 CC acid sequence or sequences grown on a Salmonella, E. coli or  
 CC Enterobacteriaceae host cell, from the host cell and introducing the  
 CC polymer into the animal in conjunction with a carrier or diluent. (1) is  
 CC useful for the expression of recombinant Agfa protein which is useful for  
 CC eliciting an immune response in an animal. In a fimbrial presentation  
 CC system the heterologous antigens are presented in high numbers (up to  
 CC 500,000 copies/cell), the hybrid fimbria protein possesses both the  
 CC immunogenicity and adhesion properties relevant for an efficient live  
 CC vaccine, the carrier fimbrial subunit proteins are usually strong  
 CC immunogens, which may be important for directing an immune response  
 CC against the inserted epitope, and hybrid fimbriae are easy and  
 CC inexpensive to purify in large amount. The present sequence is given in  
 CC the exemplification of the present invention

XX Sequence 151 AA;

Query Match 65.7%; Score 509; DB 3; Length 151;  
 Best Local Similarity 65.2%; Pred. No. 1.8e-41;  
 Matches 103; Conservative 15; Mismatches 33; Indels 0; Gaps 0;  
 QY 1 MKLLKVAFAAIVVSGSALAGVVPQWGGGNGGNGSGSPDSTLSIYQYGSANAALALQ 60  
 Db 1 MKLLKVAFAAIVVSGSALAGVVPQWGGGNGGNGSGSPDSTLSIYQYGSANAALALQ 60  
 QY 61 SDARKSETTITQSGYNGADYDQLVTVVTHEMAHAFRNATIDQWNAKNSDITVGOYGG 120  
 Db 61 TDARNSDLTTIQGGGNGADVGGGSDSSIDLTFQGFNGSATLDQWNGKNSMTVKQFGG 120  
 QY 121 NNAALVNOTASDSSVMVROVGFNGNATANQY 151  
 Db 121 GNGAAVDQTASNSVNVTVQVFGNNAHAQY 151

RESULT 15  
 ABR82651  
 ID ABR82651 standard; protein; 151 AA.  
 XX  
 AC ABR82651;  
 XX  
 DT 04-DEC-2003 (first entry)  
 XX  
 DE E. coli CsgA subunit 15 kDa protein.  
 XX

KW Plasma protein; immune response; antibacterial; vaccine; gene therapy.  
 XX Escherichia coli.  
 OS  
 XX WO2003064446-A2.  
 FN  
 XX 07-AUG-2003.  
 PD  
 XX 30-JAN-2003; 2003WO-BF000943.  
 PF  
 XX 31-JAN-2002; 2002GB-00002275.  
 PR  
 XX (HANS-) HANSA MEDICAL RES AB.  
 PA  
 XX Bjoerck L, Olsen A, Wikstroem M, Herwald H;  
 PI  
 XX WPI; 2003-646136/61.  
 DR  
 XX N-PSDB; ACF36153.

XX New isolated peptide capable of binding a mammalian plasma protein,  
 PT useful in the manufacture of a medicament for the prevention and/or  
 PT treatment of a bacterial infection, such as Escherichia coli, Salmonella  
 PT or Shigella infections.

XX Disclosure; Page 41-42; 42pp; English.

XX The invention relates to an isolated peptide capable of binding a  
 CC mammalian plasma protein or of generating an immune response in a mammal  
 CC selected from sequences shown in ABR82642, ABR82648-49. The peptide or  
 CC antibody is useful for treating a bacterial infection in a human or  
 CC animal or in the manufacture of a medicament for the prophylactic  
 CC treatment of a bacterial infection, such as Escherichia coli, Salmonella  
 CC or Shigella infection. The peptide that is immobilized on a solid support  
 CC is also useful as a reagent for determining the ability of a plasma  
 CC protein to bind to bacteria. The present sequence represents an E. coli  
 CC 15 kDa protein

XX Sequence 151 AA;

Query Match 55.0%; Score 504; DB 7; Length 151;  
 Best Local Similarity 67.5%; Pred. No. 5.5e-41;  
 Matches 102; Conservative 15; Mismatches 34; Indels 0; Gaps 0;  
 QY 1 MKLLKVAFAAIVVSGSALAGVVPQWGGGNGGNGSGSPDSTLSIYQYGSANAALALQ 60  
 Db 1 MKLLKVEAIAAIVFSGSALAGVVPQYGGGNGGNGSGSPDSTLSIYQYGGNSALALQ 60  
 QY 61 SDARKSETTITQSGYNGADYDQLVTVVTHEMAHAFRNATIDQWNAKNSDITVGOYGG 120  
 Db 61 TDARNSDLTTIQGGGNGADVGGGSDSSIDLTFQGFNGSATLDQWNGKNSMTVKQFGG 120  
 QY 121 NNAALVNOTASDSSVMVROVGFNGNATANQY 151  
 Db 121 GNGAAVDQTASNSVNVTVQVFGNNAHAQY 151

Search completed: August 2, 2004, 14:48:29  
 Job time : 45.9 secs

This Page Blank (uspto)

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: August 2, 2004, 14:40:48 ; Search time 12 Seconds  
(without alignments)  
649.627 Million cell updates/sec

Title: US-09-543-407-28  
Perfect score! 775  
Sequence: 1 MKLLKVAFAAIVVSGSALA.....DSSVMVQVGFNNATANQY 151

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 389414 seqs, 51625971 residues

Total number of hits satisfying chosen parameters: 389414

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Issued Patents AA:\*  
1: /cgn2\_6/ptodata/2/iaa/5A COMB.pap:\*  
2: /cgn2\_6/ptodata/2/iaa/5B COMB.pap:\*  
3: /cgn2\_6/ptodata/2/iaa/5A COMB.pap:\*  
4: /cgn2\_6/ptodata/2/iaa/5B COMB.pap:\*  
5: /cgn2\_6/ptodata/2/iaa/PCTUS COMB.pap:\*  
6: /cgn2\_6/ptodata/2/iaa/backfiles1.pap:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	678	87.5	151	1	US-08-233-788A-59
2	498	64.3	120	1	US-08-233-788A-57
3	85	11.0	975	4	US-09-328-352-4764
4	84.5	10.9	738	3	US-08-864-038A-3
5	84	10.8	943	4	US-09-056-556-204
6	84	10.8	943	4	US-09-072-596-199
7	84	10.8	943	4	US-09-477-135A-131
8	84	10.8	943	4	US-09-072-967-204
9	82	10.6	2123	3	US-08-968-685A-10
10	81.5	10.5	558	4	US-09-252-991A-30983
11	80	10.3	339	4	US-09-252-991A-32096
12	80	10.3	1207	4	US-09-489-039A-11518
13	79.5	10.3	745	4	US-09-336-115C-6
14	79.5	10.3	1415	4	US-09-252-991A-26438
15	78.5	10.1	714	4	US-09-841-786-4
16	78.5	10.1	3241	4	US-09-841-786-1
17	78	10.1	892	4	US-09-336-447A-5
18	78	10.1	918	4	US-09-200-650B-1
19	77.5	10.0	415	4	US-09-025-769B-280
20	77.5	10.0	518	3	US-09-043-123-2
21	77	9.9	1160	3	US-08-808-599A-24
22	76.5	9.9	873	4	US-09-336-447A-13
23	75.5	9.7	211	1	US-08-276-852-34
24	75.5	9.7	211	1	US-08-133-011-16
25	75.5	9.7	211	1	US-08-322-730A-16
26	75.5	9.7	211	1	US-08-387-874-16
27	75.5	9.7	211	1	US-08-899-575-34

28	75.5	9.7	211	1	US-08-899-575-34	Sequence 34, Appl
29	75.5	9.7	211	2	US-08-383-619-16	Sequence 16, Appl
30	75.5	9.7	211	3	US-08-907-739-16	Sequence 16, Appl
31	75.5	9.7	211	4	US-09-729-597-16	Sequence 16, Appl
32	75.5	9.7	211	5	PCT-US93-08364-16	Sequence 16, Appl
33	75.5	9.7	211	5	PCT-US95-08743-34	Sequence 34, Appl
34	75.5	9.7	218	4	US-09-495-880A-42	Sequence 42, Appl
35	75.5	9.7	244	2	US-08-553-497A-22	Sequence 22, Appl
36	75.5	9.7	266	4	US-09-495-880A-26	Sequence 26, Appl
37	75.5	9.7	293	3	US-08-438-745-4	Sequence 4, Appl
38	75.5	9.7	293	3	US-08-438-745-6	Sequence 6, Appl
39	75.5	9.7	293	3	US-09-219-019-4	Sequence 4, Appl
40	75.5	9.7	293	3	US-09-219-019-6	Sequence 6, Appl
41	75.5	9.7	293	5	PCT-US94-05669A-4	Sequence 4, Appl
42	75.5	9.7	293	5	PCT-US94-05669A-6	Sequence 6, Appl
43	75.5	9.7	432	4	US-09-403-089A-1	Sequence 1, Appl
44	75.5	9.7	461	2	US-08-463-587A-26	Sequence 26, Appl
45	75.5	9.7	461	2	US-08-463-667A-4	Sequence 4, Appl

ALIGNMENTS

RESULT 1  
US-08-233-788A-59  
; Sequence 59, Application US/08233788A  
; Patent No. 5635617  
; GENERAL INFORMATION:  
; APPLICANT: Doran, James L.  
; APPLICANT: Kay, William W.  
; APPLICANT: Collinson, Karen S.  
; APPLICANT: Clouthier, Sharon C.  
; TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR DETECTION  
; TITLE OF INVENTION: OF SALMONELLA  
; NUMBER OF SEQUENCES: 61  
; CORRESPONDENCE ADDRESS:  
; ADDRESS: Seed and Berry  
; STREET: 6300 Columbia Center, 701 Fifth Avenue  
; CITY: Seattle  
; STATE: Washington  
; COUNTRY: U.S.A.  
; ZIP: 98104-7092  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patent in Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/233,788A  
; FILING DATE: 26-APR-1994  
; CLASSIFICATION: 435  
; ATTORNEY/AGENT INFORMATION:  
; NAME: King, Joshua  
; REGISTRATION NUMBER: 35,570  
; REFERENCE/DOCKET NUMBER: 920043.403C2  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (206) 622-4900  
; TELEFAX: (206) 682-6031  
; TELEX: 3723836 SEEDANBERRY  
; INFORMATION FOR SEQ ID NO: 59:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 151 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
US-08-233-788A-59

Query Match 87.5%; Score 678; DB 1; Length 151;  
Best Local Similarity 89.4%; Pred. No. 4e-61;  
Matches 135; Conservative 1; Mismatches 15; Indels 0; Gaps 0;

QY 1 MKLLKVAFAAIVVSGSALAGVFWGGGHHNGSSGPDSTLSIYQGSAAALAQ 60  
|||||

Db 1 MKLLKVAFAALVWGSALAGVVPQWGGGNGHNGSGPDSLTSIYQGSANAALALQ 60  
QY 61 SPARKSETTITOSGNGGADYDQLVTRVVTHEMAHAFRNATIDQNAKNSDITVGOYGG 120  
Db 61 SPARKSETTITOSGNGGADYDQGVGGADNSTIEUTQNGFRNATIDQNAKNSDITVGOYGG 120  
QY 121 NNAALVNOTASDSSVMVRQVGFNNATANQY 151  
Db 121 NNPALVNOTASDSSVMVRQVGFNNATANQY 151

RESULT 2  
US-08-233-788A-57  
; Sequence 57, Application US/08233788A  
; Patent No. 5635617  
; GENERAL INFORMATION:  
; APPLICANT: Doran, James L.  
; APPLICANT: Kay, William W.  
; APPLICANT: Collinson, Karen S.  
; APPLICANT: Clouthier, Sharon C.  
; TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR DETECTION  
; TITLE OF INVENTION: OF SALMONELLA  
; NUMBER OF SEQUENCES: 61  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Seed and Berry  
; STREET: 6300 Columbia Center, 701 Fifth Avenue  
; CITY: Seattle  
; STATE: Washington  
; COUNTRY: U.S.A.  
; ZIP: 98104-7092

; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patent In Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/233,788A  
; FILING DATE: 26-APR-1994  
; CLASSIFICATION: 435  
; ATTORNEY/AGENT INFORMATION:  
; NAME: King, Joshua  
; REGISTRATION NUMBER: 35,570  
; REFERENCE/DOCKET NUMBER: 920043.403C2  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (206) 622-4900  
; TELEFAX: (206) 682-6031  
; TELEX: 3723936 SEEDANBERRY  
; INFORMATION FOR SEQ ID NO: 57:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 120 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
US-08-233-788A-57

Query Match 64.3%; Score 498; DB 1; Length 120;  
Best Local Similarity 86.6%; Pred. No. 4.6e-43;  
Matches 97; Conservative 1; Mismatches 14; Indels 0; Gaps 0;

QY 22 VVPQWGGGNGHNGSGPDSLTSIYQGSANAALALQSDARKSETTITOSGNGGADY 81  
Db 1 VVPQWGGGNGHNGSGPDSLTSIYQGSANAALALQSDARKSETTITOSGNGGADY 60  
QY 82 DQLVTRVVTHEMAHAFRNATIDQNAKNSDITVGOYGGNNAALVNOTASDS 133  
Db 61 GQADNSTIELTQNGFRNATIDQNAKNSDITVGOYGGNNAALVNOTASDS 112

RESULT 3  
US-09-328-352-4764  
; Sequence 4764, Application US/09328352  
; Patent No. 6562958  
; GENERAL INFORMATION:

; APPLICANT: Gary L. Breton et al.  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO ACINETOBACTER  
; TITLE OF INVENTION: BAUMANNII FOR DIAGNOSTICS AND THERAPEUTICS  
; FILE REFERENCE: GTC99-03PA  
; CURRENT APPLICATION NUMBER: US/09/328,352  
; CURRENT FILING DATE: 1999-06-04  
; NUMBER OF SEQ ID NOS: 8252  
; SEQ ID NO 4764  
; LENGTH: 975  
; TYPE: PRT  
; ORGANISM: Acinetobacter baumannii  
US-09-328-352-4764

Query Match 11.0%; Score 85; DB 4; Length 975;  
Best Local Similarity 25.3%; Pred. No. 4.2;  
Matches 38; Conservative 18; Mismatches 52; Indels 42; Gaps 7;  
QY 15 SGSALAGVVPQWGGGNGHNGG-GNSSGPDSTLSIYQYG-----SANA 55  
Db 300 AGNIA-----SGNGEHNFGNGGDDVDITAPITGVNLNSGNSFTLIGNSSSSVNT 353  
QY 56 ALALQSDARKSETTI-----TQSGYNGGADYDQLVTRVVTHEMAHAFRNATIDQNA 108  
Db 354 APTTSTVNDNTIDNGNSGGTGGSGNGSG-DGLNGAASNGSH----NYGNGNG 408  
QY 109 KNSDIT-----VCQYGGNNAALVNOTASDS 133  
Db 409 DDVDITSPITGIFNFGNGSFLIGNSSSSS 438

RESULT 4  
US-08-864-038A-3  
; Sequence 3, Application US/08864038A  
; Patent No. 6001592  
; GENERAL INFORMATION:  
; APPLICANT: Kunio NAKASHIMA et al.  
; TITLE OF INVENTION: NOVEL POLYPEPTIDE GENE cDNA, VECTOR  
; TITLE OF INVENTION: CONTAINING SAID cDNA, HOST CELLS TRANSFORMED WITH SAID  
; TITLE OF INVENTION: VECTOR POLYPEPTIDE PRODUCED THEREBY, METHOD OF PRODUCING  
; TITLE OF INVENTION: SAID POLYPEPTIDE, DNA ENCODING SAID POLYPEPTIDE AND ANTIBODY  
; NUMBER OF SEQUENCES: 4  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: 812-5 Hirano  
; STREET: Isehinden  
; CITY: Tsu-city  
; STATE: Mie-prefecture  
; COUNTRY: JAPAN  
; ZIP: 514-01  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Diskette, 3.50 inch, 1.44 MB storage  
; COMPUTER: IBM Compatible  
; OPERATING SYSTEM: Microsoft Windows 95  
; SOFTWARE: Word Perfect 6.1  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/864,038A  
; FILING DATE: May 28, 1997  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: JP 8-184459  
; FILING DATE: 15-July-1996  
; ATTORNEY/AGENT INFORMATION:  
; NAME: C. Bruce Hamburg  
; REGISTRATION NUMBER: 22,389  
; REFERENCE/DOCKET NUMBER: F-5610  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (212)985-2340  
; TELEFAX: (212)953-7733  
; INFORMATION FOR SEQ ID NO: 3:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 738  
; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein

RESULT 7  
US-09-47

; Sequence 131, Application US/09477135A  
; Patent No. 6572865  
; GENERAL INFORMATION:  
; APPLICANT: Nano, Francis  
; TITLE OF INVENTION: Mycobacterium Tuberculosis DNA Sequences Encoding  
; TITLE OF INVENTION: immunostimulatory Peptides  
; FILE REFERENCE: 52888  
; CURRENT APPLICATION NUMBER: US/09/477,135A  
; CURRENT FILING DATE: 2000-01-03  
; PRIOR APPLICATION NUMBER: 08990823  
; PRIOR FILING DATE: 1997-12-15  
; PRIOR APPLICATION NUMBER: US 96/10375  
; PRIOR FILING DATE: 1996-06-14  
; PRIOR APPLICATION NUMBER: 60/000,254  
; PRIOR FILING DATE: 1995-06-15  
; NUMBER OF SEQ ID NOS: 169  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 131  
; LENGTH: 943  
; TYPE: PRT  
; ORGANISM: Mycobacterium tuberculosis  
; US-09-477-135A-131

Query Match 10.8%; Score 84; DB 4; Length 943;  
Best Local Similarity 24.7%; Pred. No. 5.1;  
Matches 38; Conservative 12; Mismatches 54; Indels 50; Gaps 7;

QY 16 GSALAGVVPWGGG-GNHN-GGNSGSPDSTLSIYQYGSANAALQSDARKSETTITQS 73  
DB 571 GSGNIGVFNVGSGSLGNYNIGSGN-----LGYNIGFNGV-----DYNV 610

QY 74 GYGNGADYDQLVTRVVTHEMAHAFRRNATIDQWNAKNSDITVGYQYGN-----121  
DB 611 GFNGAGDFNQ-----GFANTGNNNIGFANTGNNNIGLSDNQGFNIASGWS 660

QY 122 ---NAALVNOTASDSSVM---VRQVGFNNATAN 149  
DB 661 GTGNSGLFSGTNNVGIFNAGTNGVANSCTGN 694

RESULT 8  
US-09-072-967-204  
; Sequence 204, Application US/09072967  
; Patent No. 6592877  
; GENERAL INFORMATION:  
; APPLICANT: Reed, Steven G.  
; APPLICANT: Skeiky, Yasir A.W.  
; APPLICANT: Dillon, Davin C.  
; APPLICANT: Campos-Neto, Antonio  
; APPLICANT: Houghton, Raymond  
; APPLICANT: Vedvick, Thomas S.  
; APPLICANT: Twardzik, Daniel R.  
; APPLICANT: Lodes, Michael J.  
; APPLICANT: Hendrickson, Ronald C.  
; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR IMMUNOTHERAPY  
; TITLE OF INVENTION: AND DIAGNOSIS OF TUBERCULOSIS  
; NUMBER OF SEQUENCES: 355  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: SEED and BERRY LLP  
; STREET: 6300 Columbia Center, 701 Fifth Avenue  
; CITY: Seattle  
; STATE: Washington  
; COUNTRY: USA  
; ZIP: 98104-7092  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; OPERATING SYSTEM: IBM PC compatible  
; SOFTWARE: PatentIn Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/09/072,967  
; FILING DATE: 05-MAY-1998  
; CLASSIFICATION:

; ATTORNEY/AGENT INFORMATION:  
; NAME: Maki, David J.  
; REGISTRATION NUMBER: 31,392  
; REFERENCE/DOCKET NUMBER: 210121.411C9  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (206) 622-4900  
; TELEFAX: (206) 682-6031  
; INFORMATION FOR SEQ ID NO: 204:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 943 amino acids  
; TYPE: amino acid  
; STRANDEDNESS:  
; TOPOLOGY: linear  
; US-09-072-967-204

Query Match 10.8%; Score 84; DB 4; Length 943;  
Best Local Similarity 24.7%; Pred. No. 5.1;  
Matches 38; Conservative 12; Mismatches 54; Indels 50; Gaps 7;

QY 16 GSALAGVVPWGGG-GNHN-GGNSGSPDSTLSIYQYGSANAALQSDARKSETTITQS 73  
DB 464 GSGNIGVFNVGSGSLGNYNIGSGN-----LGYNIGFNGV-----DYNV 503

QY 74 GYGNGADYDQLVTRVVTHEMAHAFRRNATIDQWNAKNSDITVGYQYGN-----121  
DB 504 GFNGAGDFNQ-----GFANTGNNNIGFANTGNNNIGLSDNQGFNIASGWS 553

QY 122 ---NAALVNOTASDSSVM---VRQVGFNNATAN 149  
DB 554 GTGNSGLFSGTNNVGIFNAGTNGVANSCTGN 587

RESULT 9  
US-08-968-685A-10  
; Sequence 10, Application US/08968685A  
; Patent No. 6214981  
; GENERAL INFORMATION:  
; APPLICANT: TUCKER, KENNETH  
; APPLICANT: PLOSILA, LAURA  
; TITLE OF INVENTION: MORAXELLA CATARRHALIS OUTER MEMBRANE  
; TITLE OF INVENTION: PROTEIN-106 POLYPEPTIDE, GENE SEQUENCE AND USES THEREOF  
; NUMBER OF SEQUENCES: 21  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: PENNIE & EDMONDS LLP  
; STREET: 1155 Avenue of the Americas  
; CITY: New York  
; STATE: New York  
; COUNTRY: USA  
; ZIP: 10036-2711  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/968,685A  
; FILING DATE: No. 6214981ember 12, 1997  
; CLASSIFICATION:  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Baldwin, Geraldine F.  
; REGISTRATION NUMBER: 31,232  
; REFERENCE/DOCKET NUMBER: 7969-060  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (212) 790-9090  
; TELEFAX: (212) 869-8864  
; TELEX: 66141 PENNIE  
; INFORMATION FOR SEQ ID NO: 10:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 2123 amino acids  
; TYPE: amino acid  
; STRANDEDNESS:  
; TOPOLOGY: unknown  
; MOLECULE TYPE: peptide





FEATURE:  
NAME/KEY: SIGNAL  
LOCATION: (1)...(20)  
FEATURE:  
NAME/KEY: VARIANT  
LOCATION: 721  
OTHER INFORMATION: Xaa = Any Amino Acid  
US-09-336-115C-6

Query Match 10.3%; Score 79.5; DB 4; Length 745;  
Best Local Similarity 20.5%; Pred. No. 11;  
Matches 35; Conservative 27; Mismatches 70; Indels 39; Gaps 6;  
QY 8 AFAIVVSGSALAGVVPQW-----GGGNHNGGN----- 37  
DB 102 AYQAVFLAANAAGL---WNTIGYAVMCGNGTSGPGSVIFNDQPGQDSTQITCNRPE 158  
QY 38 SSGPDSTLSIYQVGSANAALALQSDARKSETTITQSGYNGADYDQLVTRVVTHEMAHAF 97  
DB 159 STPGKMSIDEFKLEAYQIIQALKNQSGPPELG-GNGTK----VSVNRYNECROTA 213  
QY 98 RNKATIDQW-NAKNSDITVQYGGNNAALVNQTSDDSSVMVRQVGFNNAT 147  
DB 214 DINGGVQPCAKNGSSSSNGSGSSQTQTATTQDGTITTTNNKAT 264

## RESULT 14

US-09-252-991A-26438  
Sequence 26438 Application US/09252991A  
Patent No. 6551795

## GENERAL INFORMATION:

APPLICANT: Marc J. Rubenfield et al.  
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS  
TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS

FILE REFERENCE: 107196.136  
CURRENT APPLICATION NUMBER: US/09/252,991A  
CURRENT FILING DATE: 1999-02-18  
PRIOR APPLICATION NUMBER: US 60/074,788  
PRIOR FILING DATE: 1998-02-18  
PRIOR APPLICATION NUMBER: US 60/094,190  
PRIOR FILING DATE: 1998-07-27  
NUMBER OF SEQ ID NOS: 33142  
SEQ ID NO 26438  
LENGTH: 1415  
TYPE: PRT  
ORGANISM: Pseudomonas aeruginosa  
US-09-252-991A-26438

Query Match 10.3%; Score 79.5; DB 4; Length 1415;  
Best Local Similarity 28.9%; Pred. No. 25;  
Matches 44; Conservative 17; Mismatches 74; Indels 17; Gaps 9;  
QY 10 AAIIVSGSALAGVVPQWGGG--NHNGG--NSSGPDSTLSIYQVGSANAALALQSDA-- 63  
DB 930 ADPAISQ-LKDRASHYGAGLVGRNRGLIRSGSQGLTSLSGHGMNLGLVGYSSAGG 988  
QY 64 -RKSETTITQSGYNGADYDQLVTRVVTHEMAHAFNNATIDQWNAKNSDITVQYVGN- 121  
DB 989 LADVSAVDVSGNGQRLYGLGLVNLVNSGIAHATASGV-----RCTDAEALGGLIGRL 1044  
QY 122 NAALVNQTA-SDSSVMV-RQVG--FGNNATAN 149  
DB 1045 NAAINNASHGDVSLQAGRYLGLIGHNQAGN 1076

## RESULT 15

US-09-841-786-4  
Sequence 4, Application US/09841786  
Patent No. 6669940  
GENERAL INFORMATION:  
APPLICANT: NAGARAJA, T. G.  
APPLICANT: STEWART, GEORGE C.  
APPLICANT: NARAYANAN, SANJEEV K.

APPLICANT: CHENGAPPA, M. M.  
TITLE OF INVENTION: RECOMBINANT FUSOBACTERIUM NECROPHORUM LEUKOTOXIN  
TITLE OF INVENTION: VACCINE AND PREPARATION THEREOF

FILE REFERENCE: 30296  
CURRENT APPLICATION NUMBER: US/09/841,786  
CURRENT FILING DATE: 2001-04-24  
PRIOR APPLICATION NUMBER: 09/558,257  
PRIOR FILING DATE: 2000-04-25  
NUMBER OF SEQ ID NOS: 15  
SOFTWARE: PatentIn Ver. 2.1  
SEQ ID NO 4  
LENGTH: 714  
TYPE: PRT  
ORGANISM: Fusobacterium necrophorum  
US-09-841-786-4

Query Match 10.1%; Score 78.5; DB 4; Length 714;  
Best Local Similarity 22.8%; Pred. No. 13;  
Matches 31; Conservative 16; Mismatches 56; Indels 33; Gaps 6;  
QY 12 IVVSGSALAGVVPQWGGG--NHNGG--SSGPDSTLSIYQVGSANAALALQSDAR 64  
DB 35 VTSSDSTFVGA--WGGSAALQWNHIGSGNSNISAGLAGAAAVNNIQSKTSALVKNSDIR 91  
QY 65 KSE-----TTITQSGYNGADY--DQLVTRVVTHEMAHAFNNAT 102  
DB 92 NANKFYNALSGGTQVAAGAGLEAVKESG-GQKSYLLGTSASINLVNNEVSAKSENNTV 150  
QY 103 IDOWNAKNSDITVQY 118  
DB 151 AGESESKMDVDVTAY 166

Search completed: August 2, 2004, 14:58:36  
Job time : 12 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: August 2, 2004, 14:54:48 ; Search time 36.8 Seconds  
(without alignments)  
1287.123 Million cell updates/sec

Title: US-09-543-407-28

Perfect score: 775

Sequence: 1 MKLLKVAFAAIVVSGSALA.....DSSVMVRQVFGNNATANYQ 151

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1291235 seqs, 313682936 residues

Total number of hits satisfying chosen parameters: 1291235

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Published Applications AA.\*

1: /cgn2\_6/ptodata/2/pubpaa/US07\_PUBCOMB.pep.\*  
2: /cgn2\_6/ptodata/2/pubpaa/PCT\_NEW\_PUB.pep.\*  
3: /cgn2\_6/ptodata/2/pubpaa/US06\_NEW\_PUB.pep.\*  
4: /cgn2\_6/ptodata/2/pubpaa/US06\_PUBCOMB.pep.\*  
5: /cgn2\_6/ptodata/2/pubpaa/US07\_NEW\_PUB.pep.\*  
6: /cgn2\_6/ptodata/2/pubpaa/PCTUS\_PUBCOMB.pep.\*  
7: /cgn2\_6/ptodata/2/pubpaa/US08\_NEW\_PUB.pep.\*  
8: /cgn2\_6/ptodata/2/pubpaa/US08\_PUBCOMB.pep.\*  
9: /cgn2\_6/ptodata/2/pubpaa/US09A\_PUBCOMB.pep.\*  
10: /cgn2\_6/ptodata/2/pubpaa/US09B\_PUBCOMB.pep.\*  
11: /cgn2\_6/ptodata/2/pubpaa/US09C\_PUBCOMB.pep.\*  
12: /cgn2\_6/ptodata/2/pubpaa/US09\_NEW\_PUB.pep.\*  
13: /cgn2\_6/ptodata/2/pubpaa/US10A\_PUBCOMB.pep.\*  
14: /cgn2\_6/ptodata/2/pubpaa/US10B\_PUBCOMB.pep.\*  
15: /cgn2\_6/ptodata/2/pubpaa/US10C\_PUBCOMB.pep.\*  
16: /cgn2\_6/ptodata/2/pubpaa/US10\_NEW\_PUB.pep.\*  
17: /cgn2\_6/ptodata/2/pubpaa/US60\_NEW\_PUB.pep.\*  
18: /cgn2\_6/ptodata/2/pubpaa/US60\_PUBCOMB.pep.\*

/ Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	506	65.3	151	12 US-09-741-873B-4	Sequence 4, Appli
2	506	65.3	151	12 US-09-741-873B-4	Sequence 4, Appli
3	428	55.2	131	12 US-09-741-873B-2	Sequence 2, Appli
4	428	55.2	131	12 US-09-741-873B-2	Sequence 2, Appli
5	95	12.3	400	16 US-10-437-963-186417	Sequence 186417,
6	92	11.9	1074	12 US-10-282-122A-50616	Sequence 50616, A
7	89	11.5	354	10 US-09-820-843A-21	Sequence 21, Appli
8	87.5	11.3	249	9 US-09-730-374-3	Sequence 3, Appli
9	87.5	11.3	249	16 US-10-704-206-3	Sequence 3, Appli
10	87.5	11.3	445	15 US-10-369-493-20638	Sequence 20638, A
11	87.5	11.3	1448	16 US-10-408-765A-998	Sequence 998, App
12	87	11.2	597	9 US-09-793-306-146	Sequence 146, App
13	86	11.1	974	12 US-10-282-122A-44999	Sequence 44999, A
14	85.5	11.0	2204	12 US-10-282-122A-64364	Sequence 64364, A
15	85	11.0	254	10 US-09-880-748-905	Sequence 905, App

16 85 11.0 254 12 US-10-293-418-905 Sequence 905, App  
17 84.5 10.9 928 12 US-10-282-122A-54647 Sequence 54647, A  
18 84.5 10.9 978 12 US-10-282-122A-54379 Sequence 54379, A  
19 84 10.8 943 9 US-09-996-634-131 Sequence 131, App  
20 84 10.8 943 10 US-09-997-183-131 Sequence 131, App  
21 84 10.8 943 10 US-09-997-181-131 Sequence 131, App  
22 84 10.8 943 14 US-10-193-003-199 Sequence 199, App  
23 84 10.8 943 14 US-10-084-843-204 Sequence 204, App  
24 84 10.8 3300 12 US-10-282-122A-64369 Sequence 64369, A  
25 83.5 10.8 154 16 US-10-437-963-162284 Sequence 162284,  
26 83.5 10.8 288 12 US-10-424-599-283675 Sequence 283675,  
27 83 10.7 613 12 US-10-282-122A-72467 Sequence 72467, A  
28 83 10.7 793 12 US-10-282-122A-50037 Sequence 50037, A  
29 82.5 10.6 145 16 US-10-437-963-147748 Sequence 147748,  
30 82.5 10.6 319 12 US-10-244-596-32 Sequence 32, Appli  
31 82.5 10.6 319 12 US-10-244-596-33 Sequence 33, Appli  
32 82.5 10.6 580 12 US-10-647-057-4 Sequence 4, Appli  
33 82 10.6 186 12 US-10-282-122A-49412 Sequence 49412, A  
34 82 10.6 369 12 US-10-425-114-56041 Sequence 56041, A  
35 82 10.6 486 12 US-10-424-599-275468 Sequence 275468,  
36 82 10.6 507 12 US-10-425-114-57763 Sequence 57763, A  
37 82 10.6 2122 9 US-09-813-314A-9 Sequence 9, Appli  
38 81.5 10.5 485 12 US-10-282-122A-54094 Sequence 54094, A  
39 81 10.5 224 9 US-09-738-626-4894 Sequence 4894, Ap  
40 81 10.5 341 12 US-10-424-599-246651 Sequence 246651,  
41 81 10.5 703 16 US-10-437-963-108981 Sequence 108981,  
42 81 10.5 812 15 US-10-369-493-9134 Sequence 9134, Ap  
43 80.5 10.4 342 16 US-10-437-963-187064 Sequence 187064,  
44 80.5 10.4 628 12 US-10-282-122A-53269 Sequence 53269, A  
45 80 10.3 245 10 US-09-880-748-1847 Sequence 1847, Ap

#### ALIGNMENTS

##### RESULT 1

US-09-741-873B-4  
; Sequence 4, Application US/09741873B  
; Publication No. US20020081722A1  
; GENERAL INFORMATION:  
; APPLICANT: Normark, Staffan  
; APPLICANT: Olsen, Arne  
; TITLE OF INVENTION: Fibronectin Binding Protein As Well As Its Preparation  
; FILE REFERENCE: 0128893-084  
; CURRENT APPLICATION NUMBER: US/09/741,873B  
; PRIOR FILING DATE: 2003-04-04  
; PRIOR FILING DATE: 1998-05-06  
; PRIOR FILING DATE: 1998-05-06  
; PRIOR APPLICATION NUMBER: US 08/978,878  
; PRIOR FILING DATE: 1997-11-26  
; PRIOR APPLICATION NUMBER: US 07/347,189  
; PRIOR FILING DATE: 1989-05-04  
; PRIOR APPLICATION NUMBER: US 07/789,437  
; PRIOR FILING DATE: 1991-11-06  
; PRIOR APPLICATION NUMBER: US 07/970,846  
; PRIOR FILING DATE: 1992-11-03  
; PRIOR APPLICATION NUMBER: US 08/187,865  
; PRIOR FILING DATE: 1994-01-28  
; PRIOR APPLICATION NUMBER: US 08/318,519  
; NUMBER OF SEQ ID NOS: 10  
; SOFTWARE: PatentIn version 3.0  
; SEQ ID NO 4  
; LENGTH: 151  
; TYPE: PRT  
; ORGANISM: Escherichia coli  
US-09-741-873B-4

Query Match 65.3%; Score 506; DB 12; Length 151;  
Best Local Similarity 67.5%; Pred. No. 2.3e-42;  
Matches 102; Conservative 16; Mismatches 33; Indels 0; Gaps 0;  
QY 1 MKLLKVAFAAIVVSGSALAGVFWGGGNGHNGSSGPDSTLSIYQVGSAAALALQ 60

Db 1 MKLLKVAATAAIVFSGSAGVVPQVGGGNGHGGGNSGPNSELNIYQGGNSALALQ 60  
QY 61 SDARKSETTITQSGYNGADYDQLVTRVVTHEMAHAFRNATIDQNAKNSDITVQYGG 120  
Db 61 TDARNESDLITQHGCGGADVGCGSDSSIDLTLQRFNGSATLDQWNGKNSMTVKQFG 120  
QY 121 NNAALVNTASDSSVMVQVFGNNATANQY 151  
Db 121 GNGAAVDQTASNSVNVTVQFGNNATAHQY 151

## RESULT 2

US-09-741-873B-4  
; Sequence 4, Application US/09741873B  
; Publication No. US20040096965A9  
; GENERAL INFORMATION:  
; APPLICANT: Normark, Staffan  
; TITLE OF INVENTION: Fibronectin Binding Protein As Well As Its Preparation  
; FILE REFERENCE: 012889-084  
; CURRENT APPLICATION NUMBER: US/09/741,873B  
; PRIOR FILING DATE: 2003-04-04  
; PRIOR APPLICATION NUMBER: SE 8801723-1  
; PRIOR FILING DATE: 1998-05-06  
; PRIOR APPLICATION NUMBER: US 08/978,878  
; PRIOR FILING DATE: 1997-11-26  
; PRIOR APPLICATION NUMBER: US 07/347,189  
; PRIOR FILING DATE: 1999-05-04  
; PRIOR APPLICATION NUMBER: US 07/789,437  
; PRIOR FILING DATE: 1991-11-06  
; PRIOR APPLICATION NUMBER: US 07/970,846  
; PRIOR FILING DATE: 1992-11-03  
; PRIOR APPLICATION NUMBER: US 08/187,865  
; PRIOR FILING DATE: 1994-01-28  
; PRIOR APPLICATION NUMBER: US 08/318,519  
; PRIOR FILING DATE: 1994-10-05  
; NUMBER OF SEQ ID NOS: 10  
; SOFTWARE: Patentin version 3.0  
; SEQ ID NO 4  
; LENGTH: 151  
; TYPE: PRT  
; ORGANISM: Escherichia coli  
US-09-741-873B-4

Query Match 65.3%; Score 506; DB 12; Length 151;  
Best Local Similarity 67.5%; Pred. No. 2.3e-42;  
Matches 102; Conservative 16; Mismatches 33; Indels 0; Gaps 0;  
QY 1 MKLLKVAATAAIVFSGSAGVVPQVGGGNGHGGGNSGPNSELNIYQGGNSALALQ 60  
Db 1 MKLLKVAATAAIVFSGSAGVVPQVGGGNGHGGGNSGPNSELNIYQGGNSALALQ 60  
QY 61 SDARKSETTITQSGYNGADYDQLVTRVVTHEMAHAFRNATIDQNAKNSDITVQYGG 120  
Db 61 TDARNESDLITQHGCGGADVGCGSDSSIDLTLQRFNGSATLDQWNGKNSMTVKQFG 120  
QY 121 NNAALVNTASDSSVMVQVFGNNATANQY 151  
Db 121 GNGAAVDQTASNSVNVTVQFGNNATAHQY 151

## RESULT 3

US-09-741-873B-2  
; Sequence 2, Application US/09741873B  
; Publication No. US20020081722A1  
; GENERAL INFORMATION:  
; APPLICANT: Normark, Staffan  
; TITLE OF INVENTION: Fibronectin Binding Protein As Well As Its Preparation  
; FILE REFERENCE: 012889-084  
; CURRENT APPLICATION NUMBER: US/09/741,873B  
; CURRENT FILING DATE: 2003-04-04

; PRIOR APPLICATION NUMBER: SE 8801723-1  
; PRIOR FILING DATE: 1998-05-06  
; PRIOR APPLICATION NUMBER: US 08/978,878  
; PRIOR FILING DATE: 1997-11-26  
; PRIOR APPLICATION NUMBER: US 07/347,189  
; PRIOR FILING DATE: 1989-05-04  
; PRIOR APPLICATION NUMBER: US 07/789,437  
; PRIOR FILING DATE: 1991-11-06  
; PRIOR APPLICATION NUMBER: US 07/970,846  
; PRIOR FILING DATE: 1992-11-03  
; PRIOR APPLICATION NUMBER: US 08/187,865  
; PRIOR FILING DATE: 1994-01-28  
; PRIOR APPLICATION NUMBER: US 08/318,519  
; PRIOR FILING DATE: 1994-10-05  
; NUMBER OF SEQ ID NOS: 10  
; SOFTWARE: Patentin version 3.0  
; SEQ ID NO 2  
; LENGTH: 131  
; TYPE: PRT  
; ORGANISM: Escherichia coli  
US-09-741-873B-2

Query Match 55.2%; Score 428; DB 12; Length 131;  
Best Local Similarity 64.1%; Pred. No. 1.1e-34;  
Matches 84; Conservative 15; Mismatches 32; Indels 0; Gaps 0;  
QY 21 GVVPQWGGGNGHGGGNSGPNSELNIYQGGNSALALQTDARNESDLITQHGCGGAD 80  
Db 1 GVVPQWGGGNGHGGGNSGPNSELNIYQGGNSALALQTDARNESDLITQHGCGGAD 60  
QY 81 YDQLVTRVVTHEMAHAFRNATIDQNAKNSDITVQYGGNSALALQTDARNESDLITQHGCGGAD 140  
Db 61 VGGSDSSIDLTLQRFNGSATLDQWNGKNSMTVKQFGCGGAAVDQTASNSVNVTVQ 120  
QY 141 GFGNNATANQY 151  
Db 121 GFGNNATAHQY 131

## RESULT 4

US-09-741-873B-2  
; Sequence 2, Application US/09741873B  
; Publication No. US20040096965A9  
; GENERAL INFORMATION:  
; APPLICANT: Normark, Staffan  
; TITLE OF INVENTION: Fibronectin Binding Protein As Well As Its Preparation  
; FILE REFERENCE: 012889-084  
; CURRENT APPLICATION NUMBER: US/09/741,873B  
; CURRENT FILING DATE: 2003-04-04  
; PRIOR APPLICATION NUMBER: SE 8801723-1  
; PRIOR FILING DATE: 1998-05-06  
; PRIOR APPLICATION NUMBER: US 08/978,878  
; PRIOR FILING DATE: 1997-11-26  
; PRIOR APPLICATION NUMBER: US 07/347,189  
; PRIOR FILING DATE: 1989-05-04  
; PRIOR APPLICATION NUMBER: US 07/789,437  
; PRIOR FILING DATE: 1991-11-06  
; PRIOR APPLICATION NUMBER: US 07/970,846  
; PRIOR FILING DATE: 1992-11-03  
; PRIOR APPLICATION NUMBER: US 08/187,865  
; PRIOR FILING DATE: 1994-01-28  
; PRIOR APPLICATION NUMBER: US 08/318,519  
; PRIOR FILING DATE: 1994-10-05  
; NUMBER OF SEQ ID NOS: 10  
; SOFTWARE: Patentin version 3.0  
; SEQ ID NO 2  
; LENGTH: 131  
; TYPE: PRT  
; ORGANISM: Escherichia coli  
US-09-741-873B-2

Query Match 55.2%; Score 428; DB 12; Length 131;

Best Local Similarity 64.1%; Pred. No. 1.1e-34;  
Matches 84; Conservative 15; Mismatches 32; Indels 0; Gaps 0;  
QY 21 GVVPQWGGGNGHGGSSGPDSTLSIYQYGSANAALALQSDARKSETTITQSYGNGAD 80  
Db 1 GVVPQYGGGNGHGGSSGNSGPNSELNTYQYGGNSALALQTDARNSLDITQHGCGNGAD 60  
QY 81 YDQVTRVVTHEMAHAFRNATIDOWNAKNSDITVQYGGNNAALVNOQTASDSSVMVRQV 140  
Db 61 VQCGSDSSSIDLQRFQFGNSATLDWNGKNSBMTYKQFGGNGAAGVQDTASNSSVNVTVQV 120  
QY 141 GFGNNATANQY 151  
Db 121 GFGNNATAHQY 131

RESULT 5  
US-10-437-963-186417  
; Sequence 186417, Application US/10437963  
; Publication No. US20040123343A1  
; GENERAL INFORMATION:  
; APPLICANT: La Rosa, Thomas J.  
; APPLICANT: Kovalic, David K.  
; APPLICANT: Zhou, Yihua  
; APPLICANT: Cao, Yongwei  
; APPLICANT: Wu, Wei  
; APPLICANT: Boukharov, Andrey A.  
; APPLICANT: Barbazuk, Brad  
; APPLICANT: Li, Ping  
; TITLE OF INVENTION: Rice Nucleic Acid Molecules and Other Molecules Associated With  
; FILE OF INVENTION: Plants and Uses Thereof for Plant Improvement  
; FILE REFERENCE: 38-21(53221)B  
; CURRENT APPLICATION NUMBER: US/10/437,963  
; CURRENT FILING DATE: 2003-05-14  
; NUMBER OF SEQ ID NOS: 204966  
; SEQ ID NO 186417  
; LENGTH: 400  
; TYPE: PRT  
; ORGANISM: Oryza sativa  
; FEATURE:  
; OTHER INFORMATION: Clone ID: PAT\_MRT4530\_83218C.1.pap  
US-10-437-963-186417

Query Match 12.3%; Score 95; DB 16; Length 400;  
Best Local Similarity 23.6%; Pred. No. 0.52;  
Matches 37; Conservative 20; Mismatches 62; Indels 38; Gaps 5;  
QY 14 VGSALAGVVPQWGGGNGHGGSSGPDSTLSIYQYGSANAALALQSDARKSETTITQS 73  
Db 252 MNGSSSSSMEQLGGGGGWSNGSGS-----FSASWAAPFMEEAAMWATAVAPP 300  
QY 74 GYGNGADYDQV-----TRVVTHEMAHAFRNATIDOWNAKNSDIT 114  
Db 301 GNGGLAGMDQVAAQQQQQVLYQDAVGVAPTLLMWHGGGAATAGAMVVDK-SVEMVDVS 359  
QY 115 VQYGGNNAALVNOQTASDSSVMVRQVGF--GNNATAN 149  
Db 360 SAVYGGATA-----TAFDLDLAVESCGMFCGGGAGN 391

RESULT 6  
US-10-282-122A-50616  
; Sequence 50616, Application US/10282122A  
; Publication No. US20040029129A1  
; GENERAL INFORMATION:  
; APPLICANT: Wang, Liangsu  
; APPLICANT: Zamudio, Carlos  
; APPLICANT: Malone, Cheryl  
; APPLICANT: Haselbeck, Robert  
; APPLICANT: Onisen, Kari  
; APPLICANT: Zyskind, Judith  
; APPLICANT: Wall, Daniel  
; APPLICANT: Trawick, John

; APPLICANT: Carr, Grant  
; APPLICANT: Yamamoto, Robert  
; APPLICANT: Forsyth, R.  
; APPLICANT: Xu, H.  
; TITLE OF INVENTION: Identification of Essential Genes in Microorganisms  
; FILE REFERENCE: ELITRA.034A  
; CURRENT APPLICATION NUMBER: US/10/282,122A  
; CURRENT FILING DATE: 2003-02-20  
; PRIOR APPLICATION NUMBER: 60/191,078  
; PRIOR FILING DATE: 2000-03-21  
; PRIOR APPLICATION NUMBER: 60/206,848  
; PRIOR FILING DATE: 2000-05-23  
; PRIOR APPLICATION NUMBER: 60/207,727  
; PRIOR FILING DATE: 2000-05-26  
; PRIOR APPLICATION NUMBER: 60/230,335  
; PRIOR FILING DATE: 2000-09-06  
; PRIOR APPLICATION NUMBER: 60/230,347  
; PRIOR FILING DATE: 2000-09-09  
; PRIOR APPLICATION NUMBER: 60/242,578  
; PRIOR FILING DATE: 2000-10-23  
; PRIOR APPLICATION NUMBER: 60/253,625  
; PRIOR FILING DATE: 2000-11-27  
; PRIOR APPLICATION NUMBER: 60/257,931  
; PRIOR FILING DATE: 2000-12-22  
; PRIOR APPLICATION NUMBER: 60/267,636  
; PRIOR FILING DATE: 2001-02-09  
; PRIOR APPLICATION NUMBER: 60/269,308  
; PRIOR FILING DATE: 2001-02-16  
; Remaining Prior Application data removed - See File Wrapper or PALM.  
; NUMBER OF SEQ ID NOS: 78614  
; SOFTWARE: PatentIn version 3.1  
; SEQ ID NO 50616  
; LENGTH: 1074  
; TYPE: PRT  
; ORGANISM: Burkholderia mallei  
US-10-282-122A-50616

Query Match 11.9%; Score 92; DB 12; Length 1074;  
Best Local Similarity 25.1%; Pred. No. 3.5;  
Matches 39; Conservative 21; Mismatches 69; Indels 10; Gaps 5;  
QY 13 VGSALAGVVPQWGGGNGHGGSSGPDSTLSIYQYGSANAALALQSDARKSETTITQ 72  
Db 22 IASGLLIA--APAWADTVSPSGTDNVYGVDPGVTNQCNTAYGAQAKVTCGYNTA 79  
QY 73 SGYNGGADYDQVTRVVTHEMAHAFRNATIDOWNAKNS---DITVQYGGNNAALVNOQT 129  
Db 80 IGYQAGQNVN-VIDTVSIGKQATASANDALTAIGTNTKASGPADIYMLNAGAGAG--STT 136  
QY 130 ASDSSVM--VRQVCFGNA 146  
Db 137 SPDGTVTGLIRNMLGESA 155

RESULT 7  
US-09-820-843A-21  
; Sequence 21, Application US/09820843A  
; Publication No. US20030039963A1  
; GENERAL INFORMATION:  
; APPLICANT: Council of Scientific and Industrial Research  
; TITLE OF INVENTION: A COMPUTATIONAL METHOD FOR THE IDENTIFICATION OF CANDIDATE PROTEIN  
; FILE REFERENCE: Q63915  
; CURRENT APPLICATION NUMBER: US/09/820,843A  
; CURRENT FILING DATE: 2001-03-30  
; NUMBER OF SEQ ID NOS: 118  
; SOFTWARE: PatentIn version 3.0  
; SEQ ID NO 21  
; LENGTH: 354  
; TYPE: PRT  
; ORGANISM: M. tuberculosis  
; FEATURE:  
; NAME/KEY: misc\_feature

```
; OTHER INFORMATION: PPE
; NAME/KEY: misc.feature
; OTHER INFORMATION: GI11781260
US-09-820-843A-21

Query Match      11.5%; Score 89; DB 10; Length 354;
Best Local Similarity 24.1%; Pred. No. 1.7;
Matches 38; Conservative 17; Mismatches 63; Indels 40; Gaps 7;

QY 10 AAVVSGSALAG-VVPQWGGGNGHNGG-----NNSGPDSTLSIYQGSAN 54
DB 200 AGDVNSGVGNAGDVNTGLGNSGINTGPNPGLTNGTFPSAMTQAGPNS--GFFNAGTGN 257
QY 55 AALALQSDARKSETTITQSGYNGADYDQLVTRVVTHEMAFRNNATIDOWNAKNSDIT 114
DB 258 SGFGHNDPAGNSGNGIQLNSGFGNS-----CYVNTSTTSMFGNSGVLN 300
QY 115 VQGYGNNALVNQTASDSSVMVRQV---GFGNNATAN 149
DB 301 TG-YG--NSGFYNAAVNTGIFTGVWSSGFFNFGTGN 335

RESULT 8
US-09-730-374-3
; Sequence 3, Application US/09730374
; Patent No. US20010031261A1
; GENERAL INFORMATION:
; APPLICANT: Lust, John A.
; APPLICANT: Donovan, Kathleen A.
; TITLE OF INVENTION: USE OF GENETICALLY ENGINEERED ANTIBODIES
; FILE REFERENCE: 150.188US2
; CURRENT APPLICATION NUMBER: US/09/730,374
; PRIOR FILING DATE: 2001-05-10
; PRIOR APPLICATION NUMBER: PCT/US99/12512
; PRIOR FILING DATE: 1999-06-04
; PRIOR APPLICATION NUMBER: 60/088,277
; PRIOR FILING DATE: 1998-08-05
; NUMBER OF SEQ ID NOS: 4
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 3
; LENGTH: 249
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: A polypeptide encoded by an open reading frame of
; OTHER INFORMATION: SEQ ID NO:1
US-09-730-374-3

Query Match      11.3%; Score 87.5; DB 9; Length 249;
Best Local Similarity 27.3%; Pred. No. 1.6;
Matches 27; Conservative 9; Mismatches 31; Indels 31; Gaps 2;

QY 27 GGGGNGGNGGNSGSPD-----STLSIYQYGSANAAL 57
DB 126 GGGGSGGGGGGGSDIELTQSPSSFSVLGDRVTITCKASEDIYNRLAWYQKPGNAPR 185
QY 58 ALOSARKSETTITQ--SGYNGADYDQLVTRVVTHEMA 94
DB 186 LLISGATSLETGVPFRSFGSGGKDYTLTSITSLQTEDVA 224

RESULT 9
US-10-704-206-3
; Sequence 3, Application US/10704206
; Publication No. US20040141982A1
; GENERAL INFORMATION:
; APPLICANT: Lust, John A.
; APPLICANT: Donovan, Kathleen A.
; TITLE OF INVENTION: USE OF GENETICALLY ENGINEERED ANTIBODIES TO CD38 TO TREAT MULTIPLE
; FILE REFERENCE: 150.188US2
; CURRENT APPLICATION NUMBER: US/10/704,206
; CURRENT FILING DATE: 2003-11-07
```

```
; PRIOR APPLICATION NUMBER: 09/730,374
; PRIOR FILING DATE: 2000-12-05
; PRIOR APPLICATION NUMBER: PCT/US99/12512
; PRIOR FILING DATE: 1999-06-04
; PRIOR APPLICATION NUMBER: 60/088,277
; PRIOR FILING DATE: 1998-08-05
; NUMBER OF SEQ ID NOS: 5
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 3
; LENGTH: 249
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: A polypeptide encoded by an open reading frame of
; OTHER INFORMATION: SEQ ID NO:1
US-10-704-206-3

Query Match      11.3%; Score 87.5; DB 16; Length 249;
Best Local Similarity 27.3%; Pred. No. 1.6;
Matches 27; Conservative 9; Mismatches 31; Indels 31; Gaps 2;

QY 27 GGGGNGGNGGNSGSPD-----STLSIYQYGSANAAL 57
DB 126 GGGGSGGGGGGGSDIELTQSPSSFSVLGDRVTITCKASEDIYNRLAWYQKPGNAPR 185
QY 58 ALOSARKSETTITQ--SGYNGADYDQLVTRVVTHEMA 94
DB 186 LLISGATSLETGVPFRSFGSGGKDYTLTSITSLQTEDVA 224

RESULT 10
US-10-369-493-20638
; Sequence 20638, Application US/10369493
; Publication No. US20030233675A1
; GENERAL INFORMATION:
; APPLICANT: Cao, Yongwei
; APPLICANT: Hinkle, Gregory J.
; APPLICANT: Slater, Steven C.
; APPLICANT: Goldman, Barry S.
; APPLICANT: Chen, Xianfeng
; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
; FILE REFERENCE: 38-10(52052)B
; CURRENT APPLICATION NUMBER: US/10/369,493
; CURRENT FILING DATE: 2003-02-28
; PRIOR APPLICATION NUMBER: US 60/360,039
; PRIOR FILING DATE: 2002-02-21
; NUMBER OF SEQ ID NOS: 47374
; SEQ ID NO 20638
; LENGTH: 445
; TYPE: PRT
; ORGANISM: Rhodospseudomonas palustris
; FEATURE:
; NAME/KEY: unsure
; LOCATION: (1)...(445)
; OTHER INFORMATION: unsure at all Xaa locations
US-10-369-493-20638

Query Match      11.3%; Score 87.5; DB 15; Length 445;
Best Local Similarity 26.3%; Pred. No. 3.3;
Matches 45; Conservative 18; Mismatches 59; Indels 49; Gaps 8;

QY 7 AAFAA-----IVVGSALAGVVPQWGGG-----NHNGG-----GNSGPDSTLSIYQY 50
DB 19 AAFADSNVTYVLTQGTNDQANITQSGNGSVGAENGSGFLOENGLTSGA-NLLTVKQS 77
QY 51 GSANAA-----LAQSARKSETTITQSGYNGADYDQLVTRVVTHEMAFRNNATIDOW 106
DB 78 GNSNSVGRDIQKQSGAGNS-AAIFQGTGSDVELQGTG-----SNGAVPSGW 125
QY 107 NAKN-----SDITVQGYGNNALVNQTASDSSVMVRQV 141
DB 126 NWTNDPGVENKITQDSSNGSKSVIQDGNKNNVFSIKQGTNGTSTSVNQIG 176
```



Db 299 AGNIA-----SGNEHYNGNGDDVDITAPITVNLNIGNSFTLIGNSSSVNT 352  
QY 56 ALALOSDARKSETTI-----TQSGYNGADYDQLVTRVVTHEMAHAFRNNATIDOWNA 108  
Db 353 APTTNTVNDNTIDNGSGGTGSGNGSG-DGLNCAASNGEH-----NYGICNGNG 407  
QY 109 KNSDIT-----VQYGNNAALVNQTASDS 133  
Db 408 DDVDTITAPITGVNFSGNSFSLIGNSSSS 437

RESULT 14  
US-10-282-122A-64364  
; Sequence 64364, Application US/10282122A  
; Publication No. US20040029129A1  
; GENERAL INFORMATION:  
; APPLICANT: Wang, Liangsu  
; APPLICANT: Zamudio, Carlos  
; APPLICANT: Malone, Cheryl  
; APPLICANT: Haselbeck, Robert  
; APPLICANT: Ohlsen, Kari  
; APPLICANT: Zyskind, Judith  
; APPLICANT: Wall, Daniel  
; APPLICANT: Trawick, John  
; APPLICANT: Carr, Grant  
; APPLICANT: Yamamoto, Robert  
; APPLICANT: Forsyth, R.  
; APPLICANT: Xu, H.  
; TITLE OF INVENTION: Identification of Essential Genes in Microorganisms  
; FILE REFERENCE: ELITRA.034A  
; CURRENT APPLICATION NUMBER: US/10/282,122A  
; CURRENT FILING DATE: 2003-02-20  
; PRIOR FILING DATE: 2000-03-21  
; PRIOR APPLICATION NUMBER: 60/191,078  
; PRIOR FILING DATE: 2000-03-21  
; PRIOR APPLICATION NUMBER: 60/206,848  
; PRIOR FILING DATE: 2000-05-23  
; PRIOR APPLICATION NUMBER: 60/207,727  
; PRIOR FILING DATE: 2000-05-26  
; PRIOR APPLICATION NUMBER: 60/230,335  
; PRIOR FILING DATE: 2000-09-06  
; PRIOR APPLICATION NUMBER: 60/230,347  
; PRIOR FILING DATE: 2000-09-09  
; PRIOR APPLICATION NUMBER: 60/242,578  
; PRIOR FILING DATE: 2000-10-23  
; PRIOR APPLICATION NUMBER: 60/253,625  
; PRIOR FILING DATE: 2000-11-27  
; PRIOR APPLICATION NUMBER: 60/257,931  
; PRIOR FILING DATE: 2000-12-22  
; PRIOR APPLICATION NUMBER: 60/267,636  
; PRIOR FILING DATE: 2001-02-09  
; PRIOR APPLICATION NUMBER: 60/269,308  
; PRIOR FILING DATE: 2001-02-16  
; Remaining Prior Application data removed - See File Wrapper or PALM.  
; NUMBER OF SEQ ID NOS: 78614  
; SOFTWARE: PatentIn version 3.1  
; SEQ ID NO 64364  
; LENGTH: 2204  
; TYPE: PRT  
; ORGANISM: Mycobacterium tuberculosis  
US-10-282-122A-64364

Query Match 11.0%; Score 85.5; DB 12; Length 2204;  
Best Local Similarity 25.0%; Pred. No. 38;  
Matches 34; Conservative 15; Mismatches 56; Indels 31; Gaps 4;  
QY 30 GNHNGGNSGDPST-----LSIYQGSANAALALOSDARKSETTITQSGYNGADYDQLV 85  
Db 1886 GGVNIGNIGANTGTFDGLANLGSYNIGLANLGD-----DNLGFGNAGSYNIGF 1937  
QY 86 TRVVTHEMAHAFRNNATIDOWNAKNSDITVG-----QYGNNAALVNQTASDS 133  
Db 1938 ANFGDNLGFANTGYSYNI GFANTGNNNI GVGLTNGQIGISLNSGNSNIGLNFNSGSGN- 1996

QY 134 SVMVRQVFGNNTAN 149  
Db 1997 -----IGFTNSGTG 2006  
RESULT 15  
US-09-880-748-905  
; Sequence 905, Application US/09880748  
; Publication No. US20030059937A1  
; GENERAL INFORMATION:  
; APPLICANT: Ruben et al.  
; TITLE OF INVENTION: Antibodies that Immunospecifically Bind BLys  
; FILE REFERENCE: PF523  
; CURRENT APPLICATION NUMBER: US/09/880,748  
; CURRENT FILING DATE: 2001-06-15  
; PRIOR APPLICATION NUMBER: 60/212,210  
; PRIOR FILING DATE: 2000-06-15  
; PRIOR APPLICATION NUMBER: 60/240,816  
; PRIOR FILING DATE: 2000-10-17  
; PRIOR APPLICATION NUMBER: 60/276,248  
; PRIOR FILING DATE: 2001-03-16  
; PRIOR APPLICATION NUMBER: 60/277,379  
; PRIOR FILING DATE: 2001-03-21  
; PRIOR APPLICATION NUMBER: 60/293,499  
; PRIOR FILING DATE: 2001-05-25  
; NUMBER OF SEQ ID NOS: 3239  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 905  
; LENGTH: 254  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-09-880-748-905  
Query Match 11.0%; Score 85; DB 10; Length 254;  
Best Local Similarity 23.0%; Pred. No. 2.9;  
Matches 29; Conservative 11; Mismatches 42; Indels 44; Gaps 3;  
QY 13 VVSGSALAGVVPQW-----GGGNHNGGNSGDPSTLSI----- 47  
Db 104 ILTGYMGSADFQWGRGTMTVSSGGGGGGGSALEIVMTQSPGTLSPGERA 163  
QY 48 -----YQYGSANAALALOSDARKSETTITQ---SGYNGADYDQLVTV 88  
Db 164 TLSCRSQSIRSNLYAWYQKSGOAPRLLIYDVSRATGIPDRFSGSGGTDFLTISRL 223  
QY 89 VTHEMA 94  
Db 224 EPEDFA 229  
Search completed: August 2, 2004, 15:36:13  
Job time : 36.8 secs



GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: August 2, 2004, 14:48:33 ; Search time 167.9 Seconds  
(without alignments)  
877.809 Million cell updates/sec

Title: US-09-543-407-28

Perfect score: 775

Sequence: 1 MLLKVAFAAIVVSGSALA.....DSSVMVRGVGNNATANYQ 151

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 6019581 seqs, 976053577 residues

Total number of hits satisfying chosen parameters: 6019581

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

1: /cgn2\_6/ptodata/2/paa/US06 COMB.pdp.\*  
2: /cgn2\_6/ptodata/2/paa/US07 COMB.pdp.\*  
3: /cgn2\_6/ptodata/2/paa/US08 COMB.pdp.\*  
4: /cgn2\_6/ptodata/2/paa/US09 COMB.pdp.\*  
5: /cgn2\_6/ptodata/2/paa/US10 COMB.pdp.\*  
6: /cgn2\_6/ptodata/2/paa/US11 COMB.pdp.\*  
7: /cgn2\_6/ptodata/2/paa/US12 COMB.pdp.\*  
8: /cgn2\_6/ptodata/2/paa/US13 COMB.pdp.\*  
9: /cgn2\_6/ptodata/2/paa/US14 COMB.pdp.\*  
10: /cgn2\_6/ptodata/2/paa/US15 COMB.pdp.\*  
11: /cgn2\_6/ptodata/2/paa/US16 COMB.pdp.\*  
12: /cgn2\_6/ptodata/2/paa/US17 COMB.pdp.\*  
13: /cgn2\_6/ptodata/2/paa/US18 COMB.pdp.\*  
14: /cgn2\_6/ptodata/2/paa/US19 COMB.pdp.\*  
15: /cgn2\_6/ptodata/2/paa/US20 COMB.pdp.\*  
16: /cgn2\_6/ptodata/2/paa/US21 COMB.pdp.\*  
17: /cgn2\_6/ptodata/2/paa/US22 COMB.pdp.\*  
18: /cgn2\_6/ptodata/2/paa/US23 COMB.pdp.\*  
19: /cgn2\_6/ptodata/2/paa/US24 COMB.pdp.\*  
20: /cgn2\_6/ptodata/2/paa/US25 COMB.pdp.\*  
21: /cgn2\_6/ptodata/2/paa/US26 COMB.pdp.\*  
22: /cgn2\_6/ptodata/2/paa/US27 COMB.pdp.\*  
23: /cgn2\_6/ptodata/2/paa/US28 COMB.pdp.\*  
24: /cgn2\_6/ptodata/2/paa/US29 COMB.pdp.\*  
25: /cgn2\_6/ptodata/2/paa/US30 COMB.pdp.\*  
26: /cgn2\_6/ptodata/2/paa/US31 COMB.pdp.\*  
27: /cgn2\_6/ptodata/2/paa/US32 COMB.pdp.\*  
28: /cgn2\_6/ptodata/2/paa/US33 COMB.pdp.\*  
29: /cgn2\_6/ptodata/2/paa/US34 COMB.pdp.\*  
30: /cgn2\_6/ptodata/2/paa/US35 COMB.pdp.\*  
31: /cgn2\_6/ptodata/2/paa/US36 COMB.pdp.\*  
32: /cgn2\_6/ptodata/2/paa/US37 COMB.pdp.\*  
33: /cgn2\_6/ptodata/2/paa/US38 COMB.pdp.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
------------	-------	-------------	--------	----	-------------

1	775	100.0	151	19	US-09-543-407-28	Sequence 28, Appl
2	712	91.9	151	19	US-09-543-407-22	Sequence 22, Appl
3	683	88.1	151	19	US-09-543-407-5	Sequence 5, Appl
4	678	87.5	151	6	US-08-233-642A-57	Sequence 57, Appl
5	662	85.4	151	19	US-09-543-407-20	Sequence 20, Appl
6	617	79.6	151	19	US-09-543-407-26	Sequence 26, Appl
7	613	79.1	151	19	US-09-543-407-30	Sequence 30, Appl
8	605	78.1	151	19	US-09-543-407-12	Sequence 12, Appl
9	603	77.8	151	19	US-09-543-407-14	Sequence 14, Appl
10	601	77.5	151	19	US-09-543-407-18	Sequence 18, Appl
11	600	77.4	151	19	US-09-543-407-24	Sequence 24, Appl
12	596	76.9	151	19	US-09-543-407-31	Sequence 31, Appl
13	568	73.3	151	19	US-09-543-407-16	Sequence 16, Appl
14	509	65.7	151	19	US-09-543-407-7	Sequence 7, Appl
15	506	65.3	151	13	US-08-978-878-4	Sequence 4, Appl
16	506	65.3	151	21	US-09-741-873B-4	Sequence 4, Appl
17	504	65.0	151	33	US-60-352-946-2	Sequence 2, Appl
18	504	65.0	151	33	US-60-444-371-2	Sequence 2, Appl
19	498	64.3	120	6	US-08-233-642A-55	Sequence 55, Appl
20	464	59.9	109	19	US-09-543-407-34	Sequence 34, Appl
21	462	59.6	158	16	US-09-252-691-5834	Sequence 5834, Ap
22	462	59.6	158	16	US-09-252-691C-5834	Sequence 5834, Ap
23	462	59.6	158	30	US-10-417-886-5834	Sequence 2, Appl
24	428	55.2	131	13	US-08-978-878-2	Sequence 2, Appl
25	428	55.2	131	21	US-09-741-873B-2	Sequence 2, Appl
26	324	41.8	109	19	US-09-543-407-35	Sequence 35, Appl
27	257	33.2	68	19	US-09-543-407-37	Sequence 37, Appl
28	181	23.4	48	19	US-09-543-407-39	Sequence 39, Appl
29	160.5	20.7	70	19	US-09-543-407-32	Sequence 32, Appl
30	102.5	13.2	186	16	US-09-252-691-5833	Sequence 5833, Ap
31	102.5	13.2	186	16	US-09-252-691C-5833	Sequence 5833, Ap
32	102.5	13.2	186	30	US-10-417-886-5833	Sequence 5833, Ap
33	95	12.3	400	21	US-09-733-089-18734	Sequence 18734, A
34	95	12.3	400	23	US-09-816-660-18734	Sequence 18734, A
35	95	12.3	400	27	US-10-155-881-28424	Sequence 28424, A
36	95	12.3	400	30	US-10-437-963-186417	Sequence 186417, A
37	95	12.3	400	30	US-10-438-246-17264	Sequence 17264, A
38	95	12.3	1060	21	US-09-733-089-18735	Sequence 18735, A
39	95	12.3	1060	23	US-09-816-660-18735	Sequence 18735, A
40	94.5	12.2	151	19	US-09-543-407-6	Sequence 6, Appl
41	92.5	11.9	644	22	US-09-791-537-41783	Sequence 41783, A
42	92	11.9	598	22	US-09-791-537-108562	Sequence 108562, A
43	92	11.8	1074	28	US-10-382-122A-50616	Sequence 50616, A
44	91.5	11.8	478	22	US-09-791-537-23883	Sequence 23883, A
45	91.5	11.8	602	22	US-09-791-537-29231	Sequence 29231, A

ALIGNMENTS

RESULT 1

US-09-543-407-28  
; Sequence 28, Application US/09543407  
; GENERAL INFORMATION:  
; APPLICANT: White, Aaron P.  
; APPLICANT: Doran, James L.  
; APPLICANT: Collinson, S. Karen  
; APPLICANT: Kay, William W.  
; TITLE OF INVENTION: BACTERIAL FIMBRIAL SYSTEM FOR  
; FILE REFERENCE: 920043.406  
; CURRENT APPLICATION NUMBER: US/09/543,407  
; CURRENT FILING DATE: 2000-04-05  
; NUMBER OF SEQ ID NOS: 59  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 28  
; LENGTH: 151  
; TYPE: PRT  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: Recombinant Salmonella enteritidis 3b afga  
; OTHER INFORMATION: sequence containing the replacement fragment  
; OTHER INFORMATION: encoding PT3 from GP63 of Leishmania major.

US-09-543-407-28

Query Match 100.0%; Score 775; DB 19; Length 151;  
Best Local Similarity 100.0%; Pred. No. 1.2e-74;  
Matches 151; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MKLLKVAFAAIVVSGSALAGVVPQWGGGNGHNGGSGPDSITLSIYQYGSANAALQ 60  
DB 1 MKLLKVAFAAIVVSGSALAGVVPQWGGGNGHNGGSGPDSITLSIYQYGSANAALQ 60  
QY 61 SDARKSETTITQSGYNGADYDQLVTRVTHEMAHAFRNATIDOWNAKNSDITVQYGG 120  
DB 61 SDARKSETTITQSGYNGADYDQLVTRVTHEMAHAFRNATIDOWNAKNSDITVQYGG 120  
QY 121 NNAALVNQTASDSSVMVRQVGFNNATANQY 151  
DB 121 NNAALVNQTASDSSVMVRQVGFNNATANQY 151

RESULT 2

US-09-543-407-22  
; Sequence 22, Application US/09543407  
; GENERAL INFORMATION:  
; APPLICANT: White, Aaron P.  
; APPLICANT: Doran, James L.  
; APPLICANT: Collinson, S. Karen  
; APPLICANT: Kay, William W.  
; TITLE OF INVENTION: BACTERIAL FIBRIAL SYSTEM FOR  
; FILE REFERENCE: 920043.406  
; CURRENT APPLICATION NUMBER: US/09/543,407  
; CURRENT FILING DATE: 2000-04-05  
; NUMBER OF SEQ ID NOS: 59  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 22  
; LENGTH: 151  
; TYPE: PRT  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: Recombinant Salmonella enteritidis 3b afga  
; OTHER INFORMATION: sequence containing the replacement fragment  
; OTHER INFORMATION: encoding PT3 from Gp63 of Leishmania major.  
US-09-543-407-22

Query Match 91.9%; Score 712; DB 19; Length 151;  
Best Local Similarity 91.1%; Pred. No. 7.2e-68;  
Matches 144; Conservative 0; Mismatches 0; Indels 14; Gaps 2;  
QY 1 MKLLKVAFAAIVVSGSALAGVVPQWGGGNGHNGGSGPDSITLSIYQYGSANAALQ 60  
DB 1 MKLLKVAFAAIVVSGSALAGVVPQWGGGNGHNGGSGPDSITLSIYQYGSANAALQ 60  
QY 61 SDARKSETTITQSGYNGAD-----YDQLVTRVTHEMAHAFRNATIDOWNAKNSDI 113  
DB 61 SDARKSETTITQSGYNGADYDQGVGADNYDQLVTRVTHEMAHA-----DOWNAKNSDI 113  
QY 114 TVQYGGNNAALVNQTASDSSVMVRQVGFNNATANQY 151  
DB 114 TVQYGGNNAALVNQTASDSSVMVRQVGFNNATANQY 151

RESULT 3

US-09-543-407-5  
; Sequence 5, Application US/09543407  
; GENERAL INFORMATION:  
; APPLICANT: White, Aaron P.  
; APPLICANT: Doran, James L.  
; APPLICANT: Collinson, S. Karen  
; APPLICANT: Kay, William W.  
; TITLE OF INVENTION: BACTERIAL FIBRIAL SYSTEM FOR  
; FILE REFERENCE: 920043.406  
; CURRENT APPLICATION NUMBER: US/09/543,407

; CURRENT FILING DATE: 2000-04-05  
; NUMBER OF SEQ ID NOS: 59  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 5  
; LENGTH: 151  
; TYPE: PRT  
; ORGANISM: Salmonella enteritidis  
US-09-543-407-5

Query Match 88.1%; Score 683; DB 19; Length 151;  
Best Local Similarity 90.1%; Pred. No. 9.5e-65;  
Matches 136; Conservative 1; Mismatches 14; Indels 0; Gaps 0;  
QY 1 MKLLKVAFAAIVVSGSALAGVVPQWGGGNGHNGGSGPDSITLSIYQYGSANAALQ 60  
DB 1 MKLLKVAFAAIVVSGSALAGVVPQWGGGNGHNGGSGPDSITLSIYQYGSANAALQ 60  
QY 61 SDARKSETTITQSGYNGADYDQLVTRVTHEMAHAFRNATIDOWNAKNSDITVQYGG 120  
DB 61 SDARKSETTITQSGYNGADYDQGVGADNSTITELTQNGFRNATIDOWNAKNSDITVQYGG 120  
QY 121 NNAALVNQTASDSSVMVRQVGFNNATANQY 151  
DB 121 NNAALVNQTASDSSVMVRQVGFNNATANQY 151

RESULT 4

US-08-233-642A-57  
; Sequence 57, Application US/08233642A  
; GENERAL INFORMATION:  
; APPLICANT: Kay, William W.  
; APPLICANT: Collinson, S. Karen  
; APPLICANT: Clouthier, Sharon C.  
; APPLICANT: Doran, James L.  
; TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR SALMONELLA-  
; NUMBER OF SEQUENCES: 58  
; NUMBER OF SEQUENCES: -  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Seed and Berry  
; STREET: 6300 Columbia Center, 701 Fifth Avenue  
; CITY: Seattle  
; STATE: Washington  
; COUNTRY: U.S.A.  
; ZIP: 98104-7092  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/233,642A  
; FILING DATE: 26-APR-1994  
; CLASSIFICATION: 424  
; ATTORNEY/AGENT INFORMATION:  
; NAME: King, Joshua  
; REGISTRATION NUMBER: 35,570  
; REFERENCE/DOCKET NUMBER: 920043.403C3  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (206) 622-4900  
; TELEFAX: (206) 682-6031  
; TELEX: 3723836 SEEDANBERRY  
; INFORMATION FOR SEQ ID NO: 57:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 151 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
US-08-233-642A-57

Query Match 87.5%; Score 678; DB 6; Length 151;  
Best Local Similarity 89.4%; Pred. No. 3.3e-64;  
Matches 135; Conservative 1; Mismatches 15; Indels 0; Gaps 0;

QY 1 MKLLKVAFAAIVVSGSALAGVVPQWGGGNGHNGSGSPDSTLSIYQYGSANAALALQ 60  
DB 1 MKLLKVAFAAIVVSGSALAGVVPQWGGGNGHNGSGSPDSTLSIYQYGSANAALALQ 60  
QY 61 SDARKSETTTTQSGYNGADYDQLVTRVVTHEMAHAFRNNATIDQWNAKNSDITVQYGG 120  
DB 61 SDARKSETTTTQSGYNGADYDQLVTRVVTHEMAHAFRNNATIDQWNAKNSDITVQYGG 120  
QY 121 NNAALVNQTSASDSSVMVRQVGFNNATANQY 151  
DB 121 NNPAALVNQTSASDSSVMVRQVGFNNATANQY 151

RESULT 5  
US-09-543-407-20  
; Sequence 20, Application US/09543407  
; GENERAL INFORMATION:  
; APPLICANT: White, Aaron P.  
; APPLICANT: Doran, James L.  
; APPLICANT: Collinson, S. Karen  
; APPLICANT: Kay, William W.  
; TITLE OF INVENTION: BACTERIAL FIMBRIAL SYSTEM FOR  
; TITLE OF INVENTION: PRESENTATION OF HETEROLOGOUS PEPTIDE SEQUENCES  
; FILE REFERENCE: 920043.406  
; CURRENT APPLICATION NUMBER: US/09/543,407  
; CURRENT FILING DATE: 2000-04-05  
; NUMBER OF SEQ ID NOS: 59  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 20  
; LENGTH: 151  
; TYPE: PRT  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: Recombinant Salmonella enteritidis 3b afga  
; OTHER INFORMATION: sequence containing the replacement fragment  
; OTHER INFORMATION: encoding PT3 from GP63 of Leishmania major.  
US-09-543-407-20

Query Match 85.4%; Score 662; DB 19; Length 151;  
Best Local Similarity 81.9%; Pred. No. 1.7e-62;  
Matches 136; Conservative 0; Mismatches 0; Indels 30; Gaps 2;  
QY 1 MKLLKVAFAAIVVSGSALAGVVPQWGGGNGHNGSGSPDSTLSIYQYGSANAALALQ 60  
DB 1 MKLLKVAFAAIVVSGSALAGVVPQWGGGNGHNGSGSPDSTLSIYQYGSANAALALQ 60  
QY 61 SDARKSETTTTQSGYNGADYDQLVTRVVTHEMAHAFRNNATIDQ 105  
DB 61 SDARKSETTTTQSGYNGADYDQLVTRVVTHEMAHAFRNNATIDQ 105  
QY 106 WNAKNSDITVQYGGNNAALVNQTSASDSSVMVRQVGFNNATANQY 151  
DB 106 WNAKNSDITVQYGGNNAALVNQTSASDSSVMVRQVGFNNATANQY 151

RESULT 6  
US-09-543-407-26  
; Sequence 26, Application US/09543407  
; GENERAL INFORMATION:  
; APPLICANT: White, Aaron P.  
; APPLICANT: Doran, James L.  
; APPLICANT: Collinson, S. Karen  
; APPLICANT: Kay, William W.  
; TITLE OF INVENTION: BACTERIAL FIMBRIAL SYSTEM FOR  
; TITLE OF INVENTION: PRESENTATION OF HETEROLOGOUS PEPTIDE SEQUENCES  
; FILE REFERENCE: 920043.406  
; CURRENT APPLICATION NUMBER: US/09/543,407  
; CURRENT FILING DATE: 2000-04-05  
; NUMBER OF SEQ ID NOS: 59  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 26  
; LENGTH: 151

; TYPE: PRT  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: Recombinant Salmonella enteritidis 3b afga  
; OTHER INFORMATION: sequence containing the replacement fragment  
; OTHER INFORMATION: encoding PT3 from GP63 of Leishmania major.  
US-09-543-407-26

Query Match 79.6%; Score 617; DB 19; Length 151;  
Best Local Similarity 73.6%; Pred. No. 1.2e-57;  
Matches 128; Conservative 0; Mismatches 0; Indels 46; Gaps 2;  
QY 1 MKLLKVAFAAIVVSGSALAGVVPQWGGGNGHNGSGSPDSTLSIYQYGSANAALALQ 60  
DB 1 MKLLKVAFAAIVVSGSALAGVVPQWGGGNGHNGSGSPDSTLSIYQYGSANAALALQ 60  
QY 61 SDARKSETTTTQSGYNGADYDQLVTRVVTHEMAHAFRNNATIDQWNAKNSDITVQYGG 151  
DB 61 SDARKSETTTTQSGYNGADYDQLVTRVVTHEMAHAFRNNATIDQWNAKNSDITVQYGG 151  
QY 98 RNNATIDQWNAKNSDITVQYGGNNAALVNQTSASDSSVMVRQVGFNNATANQY 151  
DB 98 RNNATIDQWNAKNSDITVQYGGNNAALVNQTSASDSSVMVRQVGFNNATANQY 151

RESULT 7  
US-09-543-407-30  
; Sequence 30, Application US/09543407  
; GENERAL INFORMATION:  
; APPLICANT: White, Aaron P.  
; APPLICANT: Doran, James L.  
; APPLICANT: Collinson, S. Karen  
; APPLICANT: Kay, William W.  
; TITLE OF INVENTION: BACTERIAL FIMBRIAL SYSTEM FOR  
; TITLE OF INVENTION: PRESENTATION OF HETEROLOGOUS PEPTIDE SEQUENCES  
; FILE REFERENCE: 920043.406  
; CURRENT APPLICATION NUMBER: US/09/543,407  
; CURRENT FILING DATE: 2000-04-05  
; NUMBER OF SEQ ID NOS: 59  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 30  
; LENGTH: 151  
; TYPE: PRT  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: Recombinant Salmonella enteritidis 3b afga  
; OTHER INFORMATION: sequence containing the replacement fragment  
; OTHER INFORMATION: encoding PT3 from GP63 of Leishmania major.  
US-09-543-407-30

Query Match 79.1%; Score 613; DB 19; Length 151;  
Best Local Similarity 74.6%; Pred. No. 3.2e-57;  
Matches 129; Conservative 0; Mismatches 0; Indels 44; Gaps 2;  
QY 1 MKLLKVAFAAIVVSGSALAGVVPQWGGGNGHNGSGSPDSTLSIYQYGSANAALALQ 60  
DB 1 MKLLKVAFAAIVVSGSALAGVVPQWGGGNGHNGSGSPDSTLSIYQYGSANAALALQ 60  
QY 61 SDARKSETTTTQSGYNGADYDQLVTRVVTHEMAHAFRNNATIDQWNAKNSDITVQYGG 151  
DB 61 SDARKSETTTTQSGYNGADYDQLVTRVVTHEMAHAFRNNATIDQWNAKNSDITVQYGG 151  
QY 99 NNATIDQWNAKNSDITVQYGGNNAALVNQTSASDSSVMVRQVGFNNATANQY 151  
DB 119 NNATIDQWNAKNSDITVQYGGNNAALVNQTSASDSSVMVRQVGFNNATANQY 151

RESULT 8  
US-09-543-407-12  
; Sequence 12, Application US/09543407  
; GENERAL INFORMATION:  
; APPLICANT: White, Aaron P.  
; APPLICANT: Doran, James L.

APPLICANT: Collinson, S. Karen  
APPLICANT: Kay, William W.  
TITLE OF INVENTION: BACTERIAL FIMBRIAL SYSTEM FOR  
PRESENTATION OF HETEROLOGOUS PEPTIDE SEQUENCES  
FILE REFERENCE: 920043.406  
CURRENT APPLICATION NUMBER: US/09/543,407  
CURRENT FILING DATE: 2000-04-05  
NUMBER OF SEQ ID NOS: 59  
SOFTWARE: FastSEQ for Windows Version 4.0  
SEQ ID NO 12  
LENGTH: 151  
TYPE: PRT  
ORGANISM: Artificial Sequence  
FEATURE:  
OTHER INFORMATION: Recombinant Salmonella enteritidis 3b afga  
OTHER INFORMATION: sequence containing the replacement fragment  
OTHER INFORMATION: encoding PT3 from GP63 of Leishmania major.  
US-09-543-407-12

Query Match 78.1%; Score 605; DB 19; Length 151;  
Best Local Similarity 80.1%; Pred. No. 2.3e-56;  
Matches 121; Conservative 5; Mismatches 25; Indels 0; Gaps 0;

QY 1 MKLLKVAFAAIVVSGSALAGVVPQGGGNGHNGGNSGDPSTLSIYQYGSANAALALQ 60  
DB 1 MKLLKVAFAAIVVSGSALAGVVPQGGGNGHNGGNSGDPSTLSIYQYGSANAALALQ 60  
QY 61 SDARKSETTITQSGYNGGADYDQLVTRVVTHEMAHAFRNNATIDOWNAKNSDITVQYGG 120  
DB 61 SDARKSETTITQSGYNGGADVGQADNSTIELTQNGFRNNATIDOWNAKNSDITVQYGG 120  
QY 121 NNAALVNOTASDSSVMVROVGFNNATANQY 151  
DB 121 NNAALVNDQLVTRVVTHEMAHANNATANQY 151

RESULT 9  
US-09-543-407-14  
Sequence 14, Application US/09543407  
GENERAL INFORMATION:  
APPLICANT: White, Aaron P.  
APPLICANT: Doran, James L.  
APPLICANT: Collinson, S. Karen  
APPLICANT: Kay, William W.  
TITLE OF INVENTION: BACTERIAL FIMBRIAL SYSTEM FOR  
PRESENTATION OF HETEROLOGOUS PEPTIDE SEQUENCES  
FILE REFERENCE: 920043.406  
CURRENT APPLICATION NUMBER: US/09/543,407  
CURRENT FILING DATE: 2000-04-05  
NUMBER OF SEQ ID NOS: 59  
SOFTWARE: FastSEQ for Windows Version 4.0  
SEQ ID NO 14  
LENGTH: 151  
TYPE: PRT  
ORGANISM: Artificial Sequence  
FEATURE:  
OTHER INFORMATION: Recombinant Salmonella enteritidis 3b afga  
OTHER INFORMATION: sequence containing the replacement fragment  
OTHER INFORMATION: encoding PT3 from GP63 of Leishmania major.  
US-09-543-407-14

Query Match 77.8%; Score 603; DB 19; Length 151;  
Best Local Similarity 80.8%; Pred. No. 3.8e-56;  
Matches 122; Conservative 4; Mismatches 25; Indels 0; Gaps 0;

QY 1 MKLLKVAFAAIVVSGSALAGVVPQGGGNGHNGGNSGDPSTLSIYQYGSANAALALQ 60  
DB 1 MKLLKVAFAAIVVSGSALAGVVPQGGGNGHNGGNSGDPSTLSIYQYGSANAALALQ 60  
QY 61 SDARKSETTITQSGYNGGADYDQLVTRVVTHEMAHAFRNNATIDOWNAKNSDITVQYGG 120  
DB 61 SDARKSETTITQSGYNGGADVGQADNSTIELTQNGFRNNATIDOWNAKNSDITVQYDQ 120

QY 121 NNAALVNOTASDSSVMVROVGFNNATANQY 151  
DB 121 LVTRVVTHEMAHASVMVROVGFNNATANQY 151  
RESULT 10  
US-09-543-407-18  
Sequence 18, Application US/09543407  
GENERAL INFORMATION:  
APPLICANT: White, Aaron P.  
APPLICANT: Doran, James L.  
APPLICANT: Collinson, S. Karen  
APPLICANT: Kay, William W.  
TITLE OF INVENTION: BACTERIAL FIMBRIAL SYSTEM FOR  
PRESENTATION OF HETEROLOGOUS PEPTIDE SEQUENCES  
FILE REFERENCE: 920043.406  
CURRENT APPLICATION NUMBER: US/09/543,407  
CURRENT FILING DATE: 2000-04-05  
NUMBER OF SEQ ID NOS: 59  
SOFTWARE: FastSEQ for Windows Version 4.0  
SEQ ID NO 18  
LENGTH: 151  
TYPE: PRT  
ORGANISM: Artificial Sequence  
FEATURE:  
OTHER INFORMATION: Recombinant Salmonella enteritidis 3b afga  
OTHER INFORMATION: sequence containing the replacement fragment  
OTHER INFORMATION: encoding PT3 from GP63 of Leishmania major.  
US-09-543-407-18

Query Match 77.5%; Score 601; DB 19; Length 151;  
Best Local Similarity 80.8%; Pred. No. 6.3e-56;  
Matches 122; Conservative 4; Mismatches 25; Indels 0; Gaps 0;

QY 1 MKLLKVAFAAIVVSGSALAGVVPQGGGNGHNGGNSGDPSTLSIYQYGSANAALALQ 60  
DB 1 MKLLKVAFAAIVVSGSALAGVVPQGGGNGHNGGNSGDPYDQLVTRVVTHEMAHALQ 60  
QY 61 SDARKSETTITQSGYNGGADYDQLVTRVVTHEMAHAFRNNATIDOWNAKNSDITVQYGG 120  
DB 61 SDARKSETTITQSGYNGGADVGQADNSTIELTQNGFRNNATIDOWNAKNSDITVQYGG 120  
QY 121 NNAALVNOTASDSSVMVROVGFNNATANQY 151  
DB 121 NNAALVNOTASDSSVMVROVGFNNATANQY 151

RESULT 11  
US-09-543-407-24  
Sequence 24, Application US/09543407  
GENERAL INFORMATION:  
APPLICANT: White, Aaron P.  
APPLICANT: Doran, James L.  
APPLICANT: Collinson, S. Karen  
APPLICANT: Kay, William W.  
TITLE OF INVENTION: BACTERIAL FIMBRIAL SYSTEM FOR  
PRESENTATION OF HETEROLOGOUS PEPTIDE SEQUENCES  
FILE REFERENCE: 920043.406  
CURRENT APPLICATION NUMBER: US/09/543,407  
CURRENT FILING DATE: 2000-04-05  
NUMBER OF SEQ ID NOS: 59  
SOFTWARE: FastSEQ for Windows Version 4.0  
SEQ ID NO 24  
LENGTH: 151  
TYPE: PRT  
ORGANISM: Artificial Sequence  
FEATURE:  
OTHER INFORMATION: Recombinant Salmonella enteritidis 3b afga  
OTHER INFORMATION: sequence containing the replacement fragment  
OTHER INFORMATION: encoding PT3 from GP63 of Leishmania major.  
US-09-543-407-24

Query Match 77.4%; Score 600; DB 19; Length 151;

Best Local Similarity 81.5%; Pred. No. 8.1e-56;  
 Matches 123; Conservative 3; Mismatches 25; Indels 0; Gaps 0;  
 QY 1 MLLKVAFAAIVVSGSALAGVVPQGGGNGHNGGNSGPDSTLSIYQYGSANAALALQ 60  
 DB 1 MLLKVAFAAIVVSGSALAGVVPQGGGNGHNGGNSGPDSTLSIYQYGSANAALALQ 60  
 QY 61 SDARKSETTITQSGYNGADYDQLVTRVVTHEMAHAFRNATIDOWNAKNSDITVGYGG 120  
 DB 61 SDARKSETTITQSGYNGADYDQLVTRVVTHEMAHAFRNATIDOWNAKNSDITVGYGG 120  
 QY 121 NNAALVNQTASDSSVMVROVGFGNATANQY 151  
 DB 121 HEMAHAQNQTASDSSVMVROVGFGNATANQY 151

RESULT 12  
 US-09-543-407-31  
 ; Sequence 31, Application US/09543407  
 ; GENERAL INFORMATION:  
 ; APPLICANT: White, Aaron P.  
 ; APPLICANT: Doran, James L.  
 ; APPLICANT: Collinson, S. Karen  
 ; APPLICANT: Kay, William W.  
 ; TITLE OF INVENTION: BACTERIAL FIMBRIAL SYSTEM FOR  
 ; PRESENTATION OF HETEROLOGOUS PEPTIDE SEQUENCES  
 ; FILE REFERENCE: 920043.406  
 ; CURRENT APPLICATION NUMBER: US/09/543.407  
 ; CURRENT FILING DATE: 2000-04-05  
 ; NUMBER OF SEQ ID NOS: 59  
 ; SOFTWARE: FastSeq for Windows Version 4.0  
 ; SEQ ID NO 31  
 ; LENGTH: 131  
 ; TYPE: PRT  
 ; ORGANISM: Salmonella enteritidis  
 US-09-543-407-31

Query Match 76.9%; Score 596; DB 19; Length 131;  
 Best Local Similarity 88.5%; Pred. No. 1.8e-55;  
 Matches 116; Conservative 1; Mismatches 14; Indels 0; Gaps 0;  
 QY 21 GVVPQGGGNGHNGGNSGPDSTLSIYQYGSANAALALQSDARKSETTITQSGYNGAD 80  
 DB 1 GVVPQGGGNGHNGGNSGPDSTLSIYQYGSANAALALQSDARKSETTITQSGYNGAD 60  
 QY 81 YDQLVTRVVTHEMAHAFRNATIDOWNAKNSDITVGYGNNALVNQTASDSSVMVROV 140  
 DB 61 VQGADNSTIETQNGFRNATIDOWNAKNSDITVGYGNNALVNQTASDSSVMVROV 120  
 QY 141 GFGNATANQY 151  
 DB 121 GFGNATANQY 131

RESULT 13  
 US-09-543-407-16  
 ; Sequence 16, Application US/09543407  
 ; GENERAL INFORMATION:  
 ; APPLICANT: White, Aaron P.  
 ; APPLICANT: Doran, James L.  
 ; APPLICANT: Collinson, S. Karen  
 ; APPLICANT: Kay, William W.  
 ; TITLE OF INVENTION: BACTERIAL FIMBRIAL SYSTEM FOR  
 ; PRESENTATION OF HETEROLOGOUS PEPTIDE SEQUENCES  
 ; FILE REFERENCE: 920043.406  
 ; CURRENT APPLICATION NUMBER: US/09/543.407  
 ; CURRENT FILING DATE: 2000-04-05  
 ; NUMBER OF SEQ ID NOS: 59  
 ; SOFTWARE: FastSeq for Windows Version 4.0  
 ; SEQ ID NO 16  
 ; LENGTH: 151  
 ; TYPE: PRT  
 ; ORGANISM: Artificial Sequence

FEATURE:  
 ; OTHER INFORMATION: Recombinant Salmonella enteritidis 3b afga  
 ; OTHER INFORMATION: sequence containing the replacement fragment  
 ; OTHER INFORMATION: encoding PT3 from GP63 of Leishmania major.  
 US-09-543-407-16

Query Match 73.3%; Score 568; DB 19; Length 151;  
 Best Local Similarity 80.1%; Pred. No. 2.2e-52;  
 Matches 121; Conservative 4; Mismatches 26; Indels 0; Gaps 0;  
 QY 1 MLLKVAFAAIVVSGSALAGVVPQGGGNGHNGGNSGPDSTLSIYQYGSANAALALQ 60  
 DB 1 MLLKVAFAAIVVSGSALAGVVDQLVTRVVTHEMAHAFRNATIDOWNAKNSDITVGYGG 60  
 QY 61 SDARKSETTITQSGYNGADYDQLVTRVVTHEMAHAFRNATIDOWNAKNSDITVGYGG 120  
 DB 61 SDARKSETTITQSGYNGADYDQLVTRVVTHEMAHAFRNATIDOWNAKNSDITVGYGG 120  
 QY 121 NNAALVNQTASDSSVMVROVGFGNATANQY 151  
 DB 121 NNAALVNQTASDSSVMVROVGFGNATANQY 151

RESULT 14  
 US-09-543-407-7  
 ; Sequence 7, Application US/09543407  
 ; GENERAL INFORMATION:  
 ; APPLICANT: White, Aaron P.  
 ; APPLICANT: Doran, James L.  
 ; APPLICANT: Collinson, S. Karen  
 ; APPLICANT: Kay, William W.  
 ; TITLE OF INVENTION: BACTERIAL FIMBRIAL SYSTEM FOR  
 ; PRESENTATION OF HETEROLOGOUS PEPTIDE SEQUENCES  
 ; FILE REFERENCE: 920043.406  
 ; CURRENT APPLICATION NUMBER: US/09/543.407  
 ; CURRENT FILING DATE: 2000-04-05  
 ; NUMBER OF SEQ ID NOS: 59  
 ; SOFTWARE: FastSeq for Windows Version 4.0  
 ; SEQ ID NO 7  
 ; LENGTH: 151  
 ; TYPE: PRT  
 ; ORGANISM: Escherichia coli  
 US-09-543-407-7

Query Match 65.7%; Score 509; DB 19; Length 151;  
 Best Local Similarity 68.2%; Pred. No. 5e-46;  
 Matches 103; Conservative 15; Mismatches 33; Indels 0; Gaps 0;  
 QY 1 MLLKVAFAAIVVSGSALAGVVPQGGGNGHNGGNSGPDSTLSIYQYGSANAALALQ 60  
 DB 1 MLLKVAFAAIVVSGSALAGVVPQGGGNGHNGGNSGPDSTLSIYQYGSANAALALQ 60  
 QY 61 SDARKSETTITQSGYNGADYDQLVTRVVTHEMAHAFRNATIDOWNAKNSDITVGYGG 120  
 DB 61 TDARNSDLTITQGGNGADYDQLVTRVVTHEMAHAFRNATIDOWNAKNSDITVGYGG 120  
 QY 121 NNAALVNQTASDSSVMVROVGFGNATANQY 151  
 DB 121 GNGAADVQTASDSSVMVROVGFGNATANQY 151

RESULT 15  
 US-08-978-878-4  
 ; Sequence 4, Application US/08978878  
 ; GENERAL INFORMATION:  
 ; APPLICANT: NORMARK, Staffan  
 ; APPLICANT: OLSEN, Arne  
 ; TITLE OF INVENTION: FIBRONECTIN BINDING PROTEIN AS WELL AS ITS PREPARATION  
 ; FILE REFERENCE: 012889-081  
 ; CURRENT APPLICATION NUMBER: US/08/978.878  
 ; CURRENT FILING DATE: 1997-11-26  
 ; EARLIER APPLICATION NUMBER: SE 8801723-1  
 ; EARLIER FILING DATE: 1988-05-06

```
; EARLIER APPLICATION NUMBER: US 07/347,189
; EARLIER FILING DATE: 1989-05-04
; EARLIER APPLICATION NUMBER: US 07/789,437
; EARLIER FILING DATE: 1991-11-06
; EARLIER APPLICATION NUMBER: US 07/970,846
; EARLIER FILING DATE: 1992-11-03
; EARLIER APPLICATION NUMBER: US 08/187,865
; EARLIER FILING DATE: 1994-01-28
; EARLIER APPLICATION NUMBER: US 08/318,519
; EARLIER FILING DATE: 1994-10-05
; EARLIER APPLICATION NUMBER: US 08/495,959
; EARLIER FILING DATE: 1995-06-28
; NUMBER OF SEQ ID NOS: 10
; SOFTWARE: Patentin ver. 2.0
; SEQ ID NO 4
; LENGTH: 151
; TYPE: PRT
; ORGANISM: Escherichia coli
US-08-978-878-4

Query Match      65.3%; Score 506; DB 13; Length 151;
Best Local Similarity 67.5%; Pred. No. 1e-45;
Matches 102; Conservative 16; Mismatches 33; Indels 0; Gaps 0;

QY 1 MKLLKVAAPAAIVVSGSALAGVVPQWGGGNGGNGSSGDPSTLSIYQYGSAAALALQ 60
Db 1 MKLLKVAALAAIVFSGSAGVAVVPQYGGGNGGNGGNSGPNSELNIYQYGGNSALALQ 60

QY 61 SDARKSETTITQSGYNGGADYDQLVTRVVTHEMAHAFRNATIDQNNAKNSDITVQYGG 120
Db 61 TDARNSDLTITQGGGNGGADVGQGGDDSIDLTQRFNGSATLDQNGKNSMTVKQFGG 120

QY 121 NNAALVNOTASDSSVMVROVFGNNATANQY 151
Db 121 GNGAAVDQTASNSNVNVTQVFGNNATAHQY 151
```

Search completed: August 2, 2004, 15:26:46  
Job time : 167.9 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: August 2, 2004, 14:49:38 ; Search time 17.8 Seconds  
(without alignments)

888.146 Million cell updates/sec

Title: US-09-543-407-28

Perfect score: 775

Sequence: 1 MLLKVAAPAAIVVSGSALA.....DSSVMVQVFGNNATAY 151

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 601315 seqs, 104695340 residues

Total number of hits satisfying chosen parameters: 601315

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Pending Patents AA New:

1: /cgn2\_6/ptodata/2/paa/PCT\_NEW\_COMB.pap.\*  
2: /cgn2\_6/ptodata/2/paa/US06\_NEW\_COMB.pap.\*  
3: /cgn2\_6/ptodata/2/paa/US07\_NEW\_COMB.pap.\*  
4: /cgn2\_6/ptodata/2/paa/US08\_NEW\_COMB.pap.\*  
5: /cgn2\_6/ptodata/2/paa/US09\_NEW\_COMB.pap.\*  
6: /cgn2\_6/ptodata/2/paa/US10\_NEW\_COMB.pap.\*  
7: /cgn2\_6/ptodata/2/paa/US60\_NEW\_COMB.pap.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	506	65.3	151	5	US-09-741-873C-4
2	428	55.2	131	5	US-09-741-873C-2
3	89.5	11.5	388	5	US-09-248-796A-17306
4	87.5	11.3	619	1	PCT-US04-10229-38
5	87.5	11.3	1905	1	PCT-US04-09388-9
6	84.5	10.9	234	6	US-10-767-701-45603
7	84.5	10.9	511	6	US-10-425-115-320950
8	82	10.6	276	6	US-10-425-115-339993
9	81	10.5	147	6	US-10-425-115-193207
10	81	10.5	224	6	US-10-805-394-4894
11	81	10.5	457	5	US-09-248-796A-26659
12	81	10.5	568	1	PCT-US04-21492-458
13	80	10.3	1627	6	US-10-170-2058-16659
14	79	10.2	295	6	US-10-425-115-312468
15	79	10.2	1010	1	PCT-US04-21492-402
16	78	10.1	892	5	US-09-952-267B-5
17	78	10.1	892	6	US-10-872-769-5
18	78	10.1	892	6	US-10-872-769-5
19	77.5	10.0	246	6	US-10-854-439-511
20	77.5	10.0	415	5	US-09-490-324-280
21	77.5	10.0	520	6	US-10-479-638-21
22	77	9.9	1160	6	US-10-603-150-2
23	76.5	9.9	132	6	US-10-425-115-351875
24	76.5	9.9	248	6	US-10-170-2058-3429
25	76.5	9.9	283	6	US-10-854-439-1222
26	76.5	9.9	685	5	US-09-248-796A-15216

#### ALIGNMENTS

##### RESULT 1

US-09-741-873C-4

; Sequence 4, Application US/09741873C

; GENERAL INFORMATION:

; APPLICANT: Olsen, Staffan

; TITLE OF INVENTION: Fibronectin Binding Protein As Well As Its Preparation

; FILE REFERENCE: 012889-084

; CURRENT APPLICATION NUMBER: US/09/741,873C

; CURRENT FILING DATE: 2000-12-22

; PRIOR APPLICATION NUMBER: SE 8801723-1

; PRIOR FILING DATE: 1998-05-06

; PRIOR APPLICATION NUMBER: US 08/978,878

; PRIOR FILING DATE: 1997-11-26

; PRIOR APPLICATION NUMBER: US 07/347,189

; PRIOR FILING DATE: 1989-05-04

; PRIOR APPLICATION NUMBER: US 07/789,437

; PRIOR FILING DATE: 1991-11-06

; PRIOR APPLICATION NUMBER: US 07/970,846

; PRIOR FILING DATE: 1992-11-03

; PRIOR APPLICATION NUMBER: US 08/187,865

; PRIOR FILING DATE: 1994-01-28

; PRIOR APPLICATION NUMBER: US 08/318,519

; PRIOR FILING DATE: 1994-10-05

; NUMBER OF SEQ ID NOS: 11

; SOFTWARE: PatentIn version 3.0

; SEQ ID NO 4

; LENGTH: 151

; TYPE: PRT

; ORGANISM: Escherichia coli

US-09-741-873C-4

Query Match

Best Local Similarity 65.3%; Score 506; DB 5; Length 151;

Mismatches 102; Conservative 16; Mismatches 33; Indels 0; Gaps 0;

Qy 1 MLLKVAAPAAIVVSGSALAAGVTPQGGGNGGNGGSGSPDSTLSIYQVGSNAAALAQ 60

Db 1 MLLKVAAPAAIVVSGSALAAGVTPQGGGNGGNGGSGSPDSTLSIYQVGSNAAALAQ 60

Qy 61 SDARKSETTITQSGYNGADYDQLVTRVTVTHMAHAFRNNATIDOWNAKNSDITVQYGG 120

Db 61 TDARNSDLTITQGGGNGADVGQSDSSDLTQRFGNSATLTDQNGKNSMTVQFGG 120

Qy 121 NNAALVNOTASDSSVMVTVQVFGGNATAY 151

Db 121 GNGAAVDQTASNSVNTVQVFGGNATAY 151

Qy 121 NNAALVNOTASDSSVMVTVQVFGGNATAY 151

Db 121 GNGAAVDQTASNSVNTVQVFGGNATAY 151

Qy 121 NNAALVNOTASDSSVMVTVQVFGGNATAY 151

Db 121 GNGAAVDQTASNSVNTVQVFGGNATAY 151

Qy 121 NNAALVNOTASDSSVMVTVQVFGGNATAY 151

Db 121 GNGAAVDQTASNSVNTVQVFGGNATAY 151

Qy 121 NNAALVNOTASDSSVMVTVQVFGGNATAY 151

Db 121 GNGAAVDQTASNSVNTVQVFGGNATAY 151

Qy 121 NNAALVNOTASDSSVMVTVQVFGGNATAY 151

Db 121 GNGAAVDQTASNSVNTVQVFGGNATAY 151

Qy 121 NNAALVNOTASDSSVMVTVQVFGGNATAY 151

Db 121 GNGAAVDQTASNSVNTVQVFGGNATAY 151

Qy 121 NNAALVNOTASDSSVMVTVQVFGGNATAY 151

Db 121 GNGAAVDQTASNSVNTVQVFGGNATAY 151

## RESULT 2

US-09-741-873C-2  
; Sequence 2, Application US/09741873C  
; GENERAL INFORMATION:  
; APPLICANT: Norwalk, Staffan  
; APPLICANT: Olsen, Arne  
; TITLE OF INVENTION: Fibronectin Binding Protein As Well As Its Preparation  
; CURRENT APPLICATION NUMBER: US/09741,873C  
; CURRENT FILING DATE: 2000-12-22  
; PRIOR APPLICATION NUMBER: SE 8801723-1  
; PRIOR FILING DATE: 1998-05-06  
; PRIOR APPLICATION NUMBER: US 08/978,878  
; PRIOR FILING DATE: 1997-11-26  
; PRIOR APPLICATION NUMBER: US 07/347,189  
; PRIOR FILING DATE: 1989-05-04  
; PRIOR APPLICATION NUMBER: US 07/789,437  
; PRIOR FILING DATE: 1991-11-06  
; PRIOR APPLICATION NUMBER: US 07/970,846  
; PRIOR FILING DATE: 1992-11-03  
; PRIOR APPLICATION NUMBER: US 08/187,865  
; PRIOR FILING DATE: 1994-01-28  
; PRIOR APPLICATION NUMBER: US 08/318,519  
; PRIOR FILING DATE: 1994-10-05  
; NUMBER OF SEQ ID NOS: 11  
; SOFTWARE: Patentin version 3.0  
; SEQ ID NO 2  
; LENGTH: 131  
; TYPE: PRT  
; ORGANISM: Escherichia coli  
US-09-741-873C-2

Query Match 55.2%; Score 428; DB 5; Length 131;  
Best Local Similarity 64.1%; Pred. No. 5e-31;  
Matches 84; Conservative 15; Mismatches 32; Indels 0; Gaps 0;  
QY 21 GVVPQGGGNGGNSGPDSTLSIYQYGSANAALALQSDARKSETTITQSGYNGAD 80  
DB 1 GVVPQGGGNGGNSGPDSTLSIYQYGSANAALALQSDARKSETTITQSGYNGAD 60  
QY 81 YDOLVTRVVTHEMAHAFRNNATIDQWNAKNSDITVQYGGNNAALVNOTASDSSVMVRQV 140  
DB 61 VGGGSDSSIDLTQRFNSATLDQWNGKNSMTVQVFGGNGAAYDQTASNSVNTQV 120  
QY 141 GFGNNATANQY 151  
DB 121 GFGNNATANQY 131

## RESULT 3

US-09-248-796A-17306  
; Sequence 17306, Application US/09248796A  
; GENERAL INFORMATION:  
; APPLICANT: Keith Weinstock et al  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO CANDIDA ALBICANS  
; TITLE OF INVENTION: FOR DIAGNOSTICS AND THERAPEUTICS  
; FILE REFERENCE: 107196.132  
; CURRENT APPLICATION NUMBER: US/09/248,796A  
; CURRENT FILING DATE: 1999-02-12  
; PRIOR APPLICATION NUMBER: US 60/074,725  
; PRIOR FILING DATE: 1998-02-13  
; PRIOR APPLICATION NUMBER: US 60/096,409  
; PRIOR FILING DATE: 1998-08-13  
; NUMBER OF SEQ ID NOS: 28208  
; SEQ ID NO 17306  
; LENGTH: 388  
; TYPE: PRT  
; ORGANISM: Candida albicans  
US-09-248-796A-17306

Query Match 11.5%; Score 89.5; DB 5; Length 388;  
Best Local Similarity 22.1%; Pred. No. 3.6;  
Matches 33; Conservative 17; Mismatches 44; Indels 55; Gaps 4;

QY 30 GNHGGNSGPDSTLSIYQYGSANAALALQSDARKSETTITQSGYNGADYDQLVTRVV 89  
DB 132 GNQVGSNSYSPDT-----YGSAIGTLGVQEKTPAVTGIHSGIG-AAAYPELT----- 180  
QY 90 THEMAHAFRNNATIDQWNAKNSDITVQYGGNNAALVNOTAS----- 131  
DB 181 -----NAGVTGLAKGTAPASTATYGESPSADYSKSGATGVVPATYLN 223  
QY 132 -----DSSVMVRQVFGNNATANQY 151  
DB 224 TSGAPTGLNTAGVVGAGFGDNSTSSY 252

## RESULT 4

PCT-US04-10229-38  
; Sequence 38, Application PC/TUS0410229  
; GENERAL INFORMATION:  
; APPLICANT: Diversa Corporation  
; APPLICANT: Kerovuo, Janne  
; APPLICANT: Solbak, Arne  
; APPLICANT: Gray, Kevin  
; APPLICANT: McCann, Ryan  
; APPLICANT: Purchit, Shalaka  
; APPLICANT: Gerendash, Joel  
; APPLICANT: Janssen, Giselle  
; APPLICANT: Dahed, Samun  
; TITLE OF INVENTION: PECTATE LYASES, NUCLEIC ACIDS ENCODING  
; TITLE OF INVENTION: THEM AND METHODS FOR MAKING AND USING THEM  
; FILE REFERENCE: 564462009640  
; CURRENT APPLICATION NUMBER: PCT/US04/10229  
; CURRENT FILING DATE: 2003-04-03  
; PRIOR APPLICATION NUMBER: 60/460,842  
; PRIOR FILING DATE: 2003-04-04  
; PRIOR APPLICATION NUMBER: 60/484,798  
; PRIOR FILING DATE: 2003-07-03  
; NUMBER OF SEQ ID NOS: 134  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 38  
; LENGTH: 619  
; TYPE: PRT  
; ORGANISM: Unknown  
; FEATURE:  
; OTHER INFORMATION: Obtained from an environmental sample  
; FEATURE:  
; NAME/KEY: SIGNAL  
; LOCATION: (1)...(35)  
; FEATURE:  
; NAME/KEY: DOMAIN  
; LOCATION: (36)...(387)  
; OTHER INFORMATION: Catalytic domain  
PCT-US04-10229-38

Query Match 11.3%; Score 87.5; DB 1; Length 619;  
Best Local Similarity 26.9%; Pred. No. 9.4;  
Matches 36; Conservative 10; Mismatches 43; Indels 45; Gaps 4;  
QY 3 LLKVAFAALVWGSALAGVVPQWGGGNGGNSGPDSTLSIYQYGSANAALALQ 61  
DB 18 LTAIAAVALMAGTSFAAAAT---GGFSTTDGNGVSGARSETTASTYQOINTIIANAKLD 73  
QY 62 DARKSETTITQSGYNGADYDQLVTRVVTHEMAHAFRNNATIDQWNAKNSDITVQYGGN 121  
DB 74 DAGKVTI-----GGAYPLIIT-----YTGN 93  
QY 122 NAALVNOTASDSSV 135  
DB 94 EDLSLNQIMKDHV 107

RESULT 5  
PCT-US04-09388-9  
; Sequence 9, Application PC/TUS0409388



```

; FEATURE:
; OTHER INFORMATION: Cione ID: SORBI-28MAY03-C82834_1.pep
US-10-767-701-45603

Query Match      10.9%; Score 84.5; DB 6; Length 234;
Best Local Similarity 24.2%; Pred. No. 5.5;
Matches 39; Conservative 11; Mismatches 46; Indels 65; Gaps 6;

QY      1 MKLLKVAFAAIVV--SGSALAGVVPQW---GGGNNHNGGN-----SSGPDSTL 45
Db      1 MATTTLAALCFIVLIGIGGANAARVARYVAGGGGGGGGRWRGGASRWGSGSGC 60
QY      46 SIY-----QYGSANAALALQSDARKSETTITQSGYNGADYD 82
Db      61 GXYGAGSGXAYAGGGGGGGGGGGYGGSG-----SGYSGSGYG 102
QY      83 QLVTVVTHEMAHAFRNATIDQWNAKNSDITVQYGGNNA 123
Db      103 Q-----AGGSGSNGGAYAGGAQGGGGGGGGYGGSGS 134

RESULT 7
US-10-425-115-320950
; Sequence 320950, Application US/10425115
; GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J.
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
; FILE REFERENCE: 38-21(53222)B
; CURRENT APPLICATION NUMBER: US/10/425,115
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 369326
; SEQ ID NO 320950
; LENGTH: 511
; TYPE: PR1
; ORGANISM: Zea mays
; FEATURE:
; NAME/KEY: unsure
; LOCATION: (1)..(511)
; OTHER INFORMATION: unsure at all xaa locations
; FEATURE:
; OTHER INFORMATION: Cione ID: MRT4577_55769C.1.pep
US-10-425-115-320950

Query Match      10.9%; Score 84.5; DB 6; Length 511;
Best Local Similarity 23.3%; Pred. No. 14;
Matches 44; Conservative 22; Mismatches 64; Indels 59; Gaps 8;

QY      1 MKLLKVAFAAIVVS-----GSLAGVVPQWGGG-----GNHNGSGNSGPD 42
Db      133 LRLLEERAQEAIALASCRDVLRSAGFRPEDAWAKVIPGSDGVYAWVAANYALGRLGGDEN 192
QY      43 STLSTVQYGSANAALALQSDA-----RKSETTITQ-----SGYNGADYDQLVTR 87
Db      193 KTVGIIELGASAQITFVSDEVLPPLKLSYNYTGETTYTLYTNSFLNFQNAQDSF--- 249
QY      88 VVTHEMAH---AFRNATID-----QWNAKNSDITVQYGGNNAALVNOTASDSSVWVROV 140
Db      250 ---HEMLRSRGSFKNGTLADPCAPRGYSRNEEMMLRMSGASRSTLENQ----- 294
QY      141 GFGNNATAN 149
Db      295 -YVNNGTGN 302

RESULT 8
US-10-425-115-339993
; Sequence 339993, Application US/10425115
; GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J.

```

```

RESULT 10
US-10-805-394-4894
/ Sequence 4894, Application US/10805394
/ GENERAL INFORMATION:
/ APPLICANT: NAKAGAWA, SATOSHI
/ APPLICANT: MIZOGUCHI, HIROSHI
/ APPLICANT: ANDO, SEIKO
/ APPLICANT: HAYASHI, MIKIRO
/ APPLICANT: OCHIAI, KEIKO
/ APPLICANT: YOKOI, HARUHIKO
/ APPLICANT: TATEISHI, NAKKO
/ APPLICANT: SENOH, AKIHIRO
/ APPLICANT: IKEDA, MASATO
/ APPLICANT: OGAKI, AKIO
/ TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
/ FILE REFERENCES: 249-125
/ CURRENT APPLICATION NUMBER: US/10/805,394
/ CURRENT FILING DATE: 2004-03-22
/ PRIOR APPLICATION NUMBER: JP 99/377484
/ PRIOR FILING DATE: 1999-12-16

```

; PRIOR APPLICATION NUMBER: JP 00/135162  
 ; PRIOR FILING DATE: 2000-04-07  
 ; PRIOR APPLICATION NUMBER: JP 00/280988  
 ; PRIOR FILING DATE: 2000-08-03  
 ; NUMBER OF SEQ ID NOS: 7059  
 ; SOFTWARE: PatentIn ver. 3.0  
 ; SEQ ID NO 4894  
 ; LENGTH: 224  
 ; TYPE: PRT  
 ; ORGANISM: *Corynebacterium glutamicum*  
 US-10-805-394-4894

```

      16 GSALAGVVPQMGGGGNGHNGGNS-----GPDSTLSIYQYGSANAALALOSDARKSETTIT 71
      45 GTAIAGVLTKGGATVVEHIGSADSDIATINGDVVILAVPYPAVESITASHKDALAGKTVID 104
      QY 72 QSGYNGGADYDOLV-----TRVVTHEMAHAFRNN--ATI----- 103
      Ddb 105 ITNPLNFTFDSLVVVPVGSSTAETQAQLPTSRVLKAFNTNFAATLATGKVGDIITTVLV 164
      QY 104 --DOWNAKNSDITVGYQYGNNA 123
      Ddb 165 AGDDEDAKNALITDVNAGGLDA 186

RESULT 11
US-09-248-796A-26659
; Sequence 26659, Application US/09248796A
; GENERAL INFORMATION:
; APPLICANT: Keith Weinstock et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO CANDIDA ALBICANS

```

```

; TITLE OF INVENTION:  FOR DIAGNOSTICS AND THERAPEUTICS
;
; FILE REFERENCE:  107196.132
; CURRENT APPLICATION NUMBER:  US/09/248,796A
; CURRENT FILING DATE:  1999-02-12
; PRIOR APPLICATION NUMBER:  US 60/074,725
; PRIOR FILING DATE:  1998-02-13
; PRIOR APPLICATION NUMBER:  US 60/096,409
; PRIOR FILING DATE:  1998-08-13
; NUMBER OF SEQ ID NOS:  28208
; SEQ ID NO 26659
; LENGTH:  457
; TYPE:  PRT
; ORGANISM:  Candida albicans
; US-09-248-796A-26659
;
Query Match      10.5%   Score 81:   DB 5:   Length 457:

```

12; Gaps 6;

QY 15 SCSALAGVVPQGGGNGHNGGNS-----GPDSTLSIYQYGSANAALALQSDARKSET 68  
DB 289 SFEFLASLINGSGGGGNGSGNSYKNHSTTSKYFNSSSTATKLSKSIYNS 348  
QY 69 TITQSGY-----GNGADYDQL-----VTRVVTHEMAFAFRNNATIDOWNAKNSD 112  
DB 349 TTSRSLSVSSSSTDGGGANLFGLLNSVAAVSRTLAAESTLSTGTTTSDSANSNTKD 408  
QY 113 ITVQYGGNNAALVNQTSDDSSVMVRQVFGNNAT-ANQY 151  
DB 409 YS--SYSGTITFPSTTGSLS-----GDGNKLIIGNKY 439

RESULT 12  
PCT-US04-21492-458  
; Sequence 458, Application PC/TUS0421492  
; GENERAL INFORMATION:  
; APPLICANT: Diversa Corporation  
; APPLICANT: Steer, Brian  
; APPLICANT: Callen, Walter  
; APPLICANT: Healey, Shaun  
; APPLICANT: Pulliam, Derrick  
; TITLE OF INVENTION: GLUCANASES, NUCLEIC ACIDS ENCODING THEM AND METHODS FOR MAKING AN  
; FILE REFERENCE: THEM  
; FILE REFERENCE: 564462009540  
; CURRENT APPLICATION NUMBER: PCT/US04/21492  
; CURRENT FILING DATE: 2004-07-20  
; PRIOR APPLICATION NUMBER: 60/484,725  
; PRIOR FILING DATE: 2003-07-02  
; NUMBER OF SEQ ID NOS: 518  
; SOFTWARE: Patent in version 3.1  
; SEQ ID NO 458  
; LENGTH: 568  
; TYPE: PRT  
; ORGANISM: Unknown  
; FEATURE:  
; OTHER INFORMATION: Obtained from environmental sample  
; NAME/KEY: SIGNAL  
; LOCATION: (1)...(39)  
; FEATURE:  
; NAME/KEY: DOMAIN  
; LOCATION: (46)...(307)  
; OTHER INFORMATION: Cellulase (glycosyl hydrolase family 5)  
; FEATURE:  
; NAME/KEY: DOMAIN  
; LOCATION: (470)...(567)  
; OTHER INFORMATION: Cellulose binding domain  
; FEATURE:  
; NAME/KEY: DOMAIN  
; LOCATION: (385)...(413)  
; OTHER INFORMATION: Cellulose or protein binding domain  
PCT-US04-21492-458

Query Match 10.5%; Score 81; DB 1; Length 568;  
Best Local Similarity 20.7%; Pred. No. 32;  
Matches 35; Conservative 22; Mismatches 52; Indels 60; Gaps 7;  
QY 26 WGGGNGHN-----GGGNSGPDSTLSIYQYGSANAALALQSDARKSET 68  
DB 403 WGEWEGNSCVAVSTCSQPAPYGVVGGSSSGMSVAPSSSSSTVSLPSSSSSS 462  
QY 69 TITQSGNGADYDQLVTRVVTHEMAFA-----RNNATIDOWNAK----- 109  
DB 463 SSTASGDN--CEY-----IVSNWNTGFTCAIRINEGSSAINGVNSYSDGTSVTS 515  
QY 110 -----NSDITVQYGGNNAALVNQTSDDSSVMVRQVGF--GNNTAN 149  
DB 516 AWAATVSGNNPYSASNLGNGTTPQGTV-----EFGFGKNGKSSN 556

RESULT 13

US-10-170-205E-16659  
; Sequence 16659, Application US/10170205E  
; GENERAL INFORMATION:  
; APPLICANT: ADAMS, Mark  
; TITLE OF INVENTION: DEVICES, SUCH AS ARRAYS, COMPRISED OF HUMAN PROTEINS OR PROTEIN  
; TITLE OF INVENTION: CAPTURE AGENTS, AND USES THEREOF  
; FILE REFERENCE: CL001381  
; CURRENT APPLICATION NUMBER: US/10/170,205E  
; CURRENT FILING DATE: 2002-06-13  
; NUMBER OF SEQ ID NOS: 40312  
; SOFTWARE: Patent in version 3.2  
; SEQ ID NO 16659  
; LENGTH: 1627  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-10-170-205E-16659

Query Match 10.3%; Score 80; DB 6; Length 1627;  
Best Local Similarity 26.3%; Pred. No. 1.4e+02;  
Matches 41; Conservative 16; Mismatches 51; Indels 48; Gaps 9;  
QY 10 AAIYVSGSALAGVVPQ-----WGGGNGHN-----GGGNSGPDSTLSI 47  
DB 621 AAAAKSGHAWSGAANQEDKSPTWGCEPKPSQHWGDCQSRNPAMWAGGDDWADSSVLGH 680  
QY 48 YQYGSANAALALQSDARKS-----ETIT-QSGYNGADYDQLVTRVVTHEMAFAFEN-N 100  
DB 681 LGDKKNGS--GWDADSNRSGSGWMDTTRSGNSGNGNS-----TNTKAMPGTNWG 728  
QY 101 ATI-----DOWNAKNSDITVQYGGNNAALVNQTS 131  
DB 729 ETLFEPQONWASKPQDNVSNWGG--AASVKQTGT 762

RESULT 14  
US-10-425-115-312468  
; Sequence 312468, Application US/10425115  
; GENERAL INFORMATION:  
; APPLICANT: La Rosa, Thomas J.  
; APPLICANT: Kovalic, David K.  
; APPLICANT: Zhou, Yihua  
; APPLICANT: Cao, Yongwei  
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With  
; FILE REFERENCE: 38-21(53222)B  
; CURRENT APPLICATION NUMBER: US/10/425,115  
; CURRENT FILING DATE: 2003-04-28  
; NUMBER OF SEQ ID NOS: 369326  
; SEQ ID NO 312468  
; LENGTH: 295  
; TYPE: PRT  
; ORGANISM: Zea mays  
; FEATURE:  
; OTHER INFORMATION: Clone ID: MFT4577\_48027C.1.pep  
US-10-425-115-312468

Query Match 10.2%; Score 79; DB 6; Length 295;  
Best Local Similarity 25.4%; Pred. No. 23;  
Matches 32; Conservative 13; Mismatches 43; Indels 38; Gaps 5;  
QY 26 WGGGNGHNGGNSGPDSTLSIYQYGSANAALALQSDARKSETTITQSGYNGADYDQLV 85  
DB 138 YGGGYSGGGYSSG--GYAANGYGVGSGGNYSNAGGGYSGS-----DGYNGA----- 186  
QY 86 TRVVTHEMAFAFRNNATIDOWNAKNSDITVQYGGNNAALVNQTSDDSSVMVRQVFGNN 145  
DB 187 -----ASGGYANNLS--SGYNSGRYNTIGSSDNTG-----GYN 219  
QY 146 ATANQY 151  
DB 220 SYNPY 225

```
RESULT 15
PCT-US04-21492-402
; Sequence 402, Application PC/TUS0421492
; GENERAL INFORMATION:
; APPLICANT: Diversa Corporation
; APPLICANT: Steer, Brian
; APPLICANT: Callen, Walter
; APPLICANT: Healey, Shaun
; APPLICANT: Pulliam, Derrick
; TITLE OF INVENTION: GLUCANASES, NUCLEIC ACIDS ENCODING THEM AND METHODS FOR MAKING AN
; TITLE OF INVENTION: THEM
; FILE REFERENCE: 564462009540
; CURRENT APPLICATION NUMBER: PCT/US04/21492
; CURRENT FILING DATE: 2004-07-20
; PRIOR APPLICATION NUMBER: 60/484,725
; PRIOR FILING DATE: 2003-07-02
; NUMBER OF SEQ ID NOS: 518
; SOFTWARE: Patent In version 3.1
; SEQ ID NO 402
; LENGTH: 1010
; TYPE: PRT
; ORGANISM: Bacteria
; FEATURE:
; NAME/KEY: SIGNAL
; LOCATION: (1)...(30)
; FEATURE:
; NAME/KEY: DOMAIN
; LOCATION: (393)...(428)
; OTHER INFORMATION: Carbohydrate binding domain
; FEATURE:
; NAME/KEY: DOMAIN
; LOCATION: (39)...(300)
; OTHER INFORMATION: Cellulase (glycosyl hydrolase family 5)
; FEATURE:
; NAME/KEY: DOMAIN
; LOCATION: (610)...(959)
; OTHER INFORMATION: Glycosyl hydrolases family 6
; FEATURE:
; NAME/KEY: DOMAIN
; LOCATION: (493)...(521)
; OTHER INFORMATION: Cellulose or protein binding domain
PCT-US04-21492-402

Query Match 10.2%; Score 79; DB 1; Length 1010;
Best Local Similarity 20.8%; Pred. No. 96;
Matches 32; Conservative 27; Mismatches 45; Indels 50; Gaps 7;

QY 12 IVVSGSALAGVVPQWG---GGGNHGGKSGSPDSTLSIYQYGSNAALALQSDARKSET 68
Db 316 LTASGTLVKNVSNWGTITGNSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSS 367
QY 69 TITQSGVGN-----GADY-----DQVTRVVTHMAHAFRNATIDQW 106
Db 368 SSGSTGGKNCAGVYVFNWTARDWSGAYNANAGDQWV-----YQNSLYRANW 416
QY 107 ---NAKNSD---ITVGQYGNNAALVNTASDSS 134
Db 417 YTNVFPQSDASWTSLGACGNGSTTSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSS 450
```

Search completed: August 2, 2004, 15:29:55  
Job time : 17.8 secs



Db 121 NNAALVNQTASDSSVMVRQVGFNNATANQY 151

RESULT 2

AI0635

major curlin chain precursor [imported] - Salmonella enterica subsp. enterica serovar Typhimurium

C:Species: Salmonella enterica subsp. enterica serovar Typhi

A:Note: this species has also been called Salmonella typhi

C>Date: 09-Nov-2001 #sequence\_revision 09-Nov-2001 #text\_change 18-Nov-2002

C:Accession: AI0635

R:Parkhill, J.; Doughty, G.; James, K.D.; Thomson, N.R.; Pickard, D.; Wain, J.; Churcher, C.; Parkhill, J.; Cronin, A.; Davis, P.; Davies, R.M.; Dowd, L.; White, N.; Farrar, S.; Moule, S.; O'Gaora, P.

Nature 413, 848-852, 2001

A:Authors: Parry, C.; Quail, M.; Rutherford, K.; Simmonds, M.; Skelton, J.; Stevens, K.

A:Title: Complete genome sequence of a multiple drug resistant Salmonella enterica serovar Typhimurium

A:Reference number: AB0502; MUID:21534947; PMID:11677608

A:Accession: AI0635

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-151 <PAR>

A:Cross-references: GB:AL513382; PIDN:CAD08268.1; PID:gl6502315; GSPDB:GN00176

C:Genetics:

A:Gene: STY1181

Query Match 88.1%; Score 683; DB 2; Length 151;

Best Local Similarity 90.1%; Pred. No. 1.4e-51;

Matches 136; Conservative 1; Mismatches 14; Indels 0; Gaps 0;

Qy 1 MKLLKVAAPAAIVSGSALAGVVPQWGGGNGGNGSGGPDSTLSIYQYGSANAALALQ 60

Db 1 MKLLKVAAPAAIVSGSALAGVVPQWGGGNGGNGSGGPDSTLSIYQYGSANAALALQ 60

Qy 61 SDARKSETTITQSGYNGADYDQLVTRVVTHEMAFRNNATIDQWNAKNSDITVQYGG 120

Db 61 SDARKSETTITQSGYNGADYDQLVTRVVTHEMAFRNNATIDQWNAKNSDITVQYGG 120

Qy 121 NNAALVNQTASDSSVMVRQVGFNNATANQY 151

Db 121 NNAALVNQTASDSSVMVRQVGFNNATANQY 151

RESULT 3

S70788

curlin protein csgA precursor - Escherichia coli (strain K-12)

N:Alternate names: csgA protein; major curlin protein

C:Species: Escherichia coli

C>Date: 12-Feb-1998 #sequence\_revision 20-Feb-1998 #text\_change 01-Mar-2002

C:Accession: S70788; G54846; S31202; S34560; S34559

R:Hammar, M.; Arngvist, A.; Bian, Z.; Olsen, A.; Normark, S.

Mol. Microbiol. 18, 661-670, 1995

A:Title: Expression of two csg operons is required for production of fibronectin- and curli

A:Reference number: S70788; MUID:96414468; PMID:8817489

A:Accession: S70788

A:Status: nucleic acid sequence not shown; translation not shown

A:Molecule type: DNA

A:Residues: 1-151 <HAM>

A:Cross-references: EMBL:X90754; NID:g1147558; PIDN:CAA62282.1; PID:g1147564

A:Experimental source: strain K12, substrain W3110

A:Note: The nucleotide sequence was submitted to the EMBL Data Library, August 1995

R:Blattner, F.R.; Plunkett III, G.; Bloch, C.A.; Perna, N.T.; Burland, V.; Riley, M.; Crook, S.; Holm, M.; DeLong, E.P.; Olsen, G.J.; Mount, R.B.; Shao, Y.

Science 277, 1453-1462, 1997

A:Title: The complete genome sequence of Escherichia coli K-12.

A:Reference number: AG4720; MUID:97426617; PMID:9278503

A:Accession: G64846

A:Status: nucleic acid sequence not shown; translation not shown

A:Molecule type: DNA

A:Residues: 1-151 <BLAT>

A:Cross-references: GB:AE000205; GB:U00096; NID:g1787265; PIDN:AAC74126.1; PID:g1787279;

A:Experimental source: strain K-12, substrain MGL655

R:Olsen, A.; Arngvist, A.; Hammar, M.; Sukupolvi, S.; Normark, S.

Mol. Microbiol. 7, 523-536, 1993

A:Title: The RpoS sigma factor relieves H-NS-mediated transcriptional repression of csgA,

A:Reference number: S31202; MUID:93211294; PMID:8459772

A:Accession: S31202

A:Molecule type: DNA

A:Residues: 1-6,'V', 8-151 <OLSI>

A:Cross-references: EMBL:L04979

A:Accession: S34560

A:Molecule type: protein

A:Residues: 21-42;44-50 <OLS2>

R:Olsen, A.N.; Arngvist, A.M.

submitted to the EMBL Data Library, October 1992

A:Reference number: S34559

A:Accession: S34559

A:Molecule type: DNA

A:Residues: 1-133,'RORDSGWLW' <OLS3>

A:Cross-references: EMBL:L04979; NID:g290424; PIDN:AAA23616.1; PID:g290425

A:Experimental source: strain K-12, substrain W3110

C:Genetics:

A:Gene: csgA

A:Map position: 23.15

C:Function:

A:Description: major component of wild-type curli; interaction between CsgA and CsgB trig

A:Note: curli are thin, coiled fibers expressed on the surface of Escherichia coli that i

and H-kininogen; in the absence of CsgA, CsgB can self-assemble into polymers

F:1-20/Domain: signal sequence #status predicted <SIG>

F:21-151/Product: curlin #status experimental <MAT>

Query Match 65.7%; Score 509; DB 2; Length 151;

Best Local Similarity 68.2%; Pred. No. 1.1e-36;

Matches 103; Conservative 15; Mismatches 33; Indels 0; Gaps 0;

Qy 1 MKLLKVAAPAAIVSGSALAGVVPQWGGGNGGNGSGGPDSTLSIYQYGSANAALALQ 60

Db 1 MKLLKVAAPAAIVSGSALAGVVPQWGGGNGGNGSGGPDSTLSIYQYGSANAALALQ 60

Qy 61 SDARKSETTITQSGYNGADYDQLVTRVVTHEMAFRNNATIDQWNAKNSDITVQYGG 120

Db 61 TDARNSDLTITQSGGNGADYDQGGSDSDSIDITQSGFGNSATLDQWNGKNSMTVKQFGG 120

Qy 121 NNAALVNQTASDSSVMVRQVGFNNATANQY 151

Db 121 GNGRAVDQTASNSVNVTVQVGFNNATAHQY 151

RESULT 4

D90806

curlin major subunit CsgA [imported] - Escherichia coli (strain O157:H7, substrain RIMD (

C:Species: Escherichia coli

C>Date: 18-Jul-2001 #sequence\_revision 18-Jul-2001 #text\_change 18-Jul-2001

C:Accession: D90806

R:Hayaashi, T.; Makino, K.; Ohnishi, M.; Kurokawa, K.; Ishii, K.; Yokoyama, K.; Han, C.G.;

gasawara, N.; Yasunaga, T.; Kuhara, S.; Shiba, T.; Hattori, M.; Shinagawa, H.

DNA Res. 8, 11-22, 2001

A:Title: Complete genome sequence of enterohemorrhagic Escherichia coli O157:H7 and genom

A:Reference number: A99629; MUID:21156231; PMID:11258796

A:Accession: D90806

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-152 <HAY>

A:Cross-references: GB:BA000007; PIDN:BA34843.1; PID:g13360880; GSPDB:GN00154

A:Experimental source: strain O157:H7, substrain RIMD 0509952

C:Genetics:

A:Gene: ECs1420

Query Match 62.9%; Score 487.5; DB 2; Length 152;

Best Local Similarity 66.4%; Pred. No. 7.6e-35;

Matches 101; Conservative 16; Mismatches 34; Indels 1; Gaps 1;

Qy 1 MKLLKVAAPAAIVSGSALAGVVPQWGGGNGGNGSGGPDSTLSIYQYGSANAALAL 59

Db 1 MKLLKVAAPAAIVSGSALAGVVPQWGGGNGGNGSGGPDSTLSIYQYGSANAALAL 60

Qy 60 QSDARKSETTITQSGYNGADYDQLVTRVVTHEMAFRNNATIDQWNAKNSDITVQYGG 119

```

Query Match      13.8%; Score 107; DB 2; Length 2174;
Best Local Similarity 26.2%; Pred. No. 0.62;
Matches 37; Conservative 21; Mismatches 61; Indels 22; Gaps 6;

Qy      11 AIVVSGALAGVVPQ--WGGGGNHNGGSSGDPDSTLSIYQYGSAALALQSDAR---K 65
      693 AIAATGAGAVGIIIAQISGGGGN---GGNATGGDAGGFSQIGGGGGGGVANTANVCFK 749

Db      66 SETTITQSGYNGADYDQLVTRVVVTHEMAHFRNNATIDOWNAK---NSDITVQYQYGN 122
      750 GLTLITQGSAAAG-----IVAQSVGGGGTGGTASSVSAGIGFTASVAVGGTGNG 800

Qy      123 AA--LYNQATASDSSVMVRQVG 141
      801 GAGGEVSVSLTDSAIRTGQGG 821

Db

RESULT 7
S42136
C:cnjB protéin - Tetrahymena thermophila
C:Species: Tetrahymena thermophila
C:Date: 19-Mar-1997 #sequence revision 18-Jul-1997 #text_change 07-Dec-1999
C:Accession: S42136; S42135; S03650
R:Taylor, F.M.; Martindale, D.W.
submitted to the EMBL Data Library, October 1992
A:Reference number: S42136

```

A;Accession: S42136  
A;Molecule type: DNA  
A;Residues: 1-1748 <TAY>  
A;Cross-references: EMBL:L03710; NID:g161751; PID:g161752  
R;Taylor, F.M.; Martindale, D.W.  
Nucleic Acids Res. 21, 4610-4614, 1993  
A;Title: Retroviral-type zinc fingers and glycine-rich repeats in a protein encoded by the *gag* gene of the Rous sarcoma virus  
A;Reference number: S42135; MUID:94051563; PMID:8233798  
A;Accession: S42135  
A;Molecule type: DNA  
A;Residues: 1164-1174;1179-1198;1233-1252;1285-1293;1297-1309;1316-1326;1331-1340  
A;Cross-references: EMBL:L03710  
R;Martindale, D.W.; Taylor, F.M.  
Nucleic Acids Res. 15, 2189-2201, 1988  
A;Title: Multiple introns in a conjugation-specific gene from *Tetrahymena thermophila*  
A;Reference number: S03650; MUID:8819861; PMID:3357771

```

A:Accession: 503650
A:Molecule type: DNA
A:Residues: 236-250, 'I', 252-255, 'N', 257-773 <MAR>
A:Cross-references: EMBL:X06462
C:Genetics:
C:Gene: cnjB
A:Genetic code: SCS5
A:Introns: 85/3; 136/1; 157/3; 201/2; 290/2; 327/3; 499/1; 573/2; 607/3; 708/3
C:Keywords: zinc finger
F:1164-1450/Region: Glycine-rich
F:1451-1464/Region: zinc finger CCHC motif
F:1478-1491/Region: zinc finger CCHC motif
F:1501-1514/Region: zinc finger CCHC motif
F:1530-1543/Region: zinc finger CCHC motif
F:1555-1568/Region: zinc finger CCHC motif
F:1579-1592/Region: zinc finger CCHC motif
F:1602-1615/Region: zinc finger CCHC motif
F:1626-1748/Region: glycine-rich

Query March 12.5%; Score 97; DB 2; Length 1748;
Best Local Similarity 28.1%, Pred. No. 3.5;
Matches 36; Conservative 20; Mismatches 32; Indels 40; Gaps 7

QY 25 QWGGGGNHNG---GNSGGPDSTLSIYOYGSAALALQSDARKSETTITOSGYNGADY 81
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1640 QFGGGGNSGGQSGWTSGSDWN-----CQSNVQES-ITTTSGGWSGSSG 168

QY 82 DQLYTRVVTHEMAFRNRAIDQWNAKNSDITVQYGGNNAALVNOTASDSSVMVRQG 141
      :||: :||: :||: :||: :||: :||: :||: :||: :||: :||:
Db 1685 NQ-----TGGGWSGN-----DQQQQQNTGGGGWSSNS---NQTNSS----- 172

```





```

Db      365  SGTGNG-----FGSGNNNIGFFNSG-DGNTGFFNSGDGN-----TG 401

Qy      75  YNGGADYDOLVTRVYVTHEMAHAFRNATIDOWNAKNSDITVGYGNNALV---NOTAS 131
      402  FGNA:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
      -----GNINTGFWNAGNLNTFGSAGNGNVGIFDGGNSNG 441

Qy      132  DSSVMVRQVGFEGNNATAN 149
      442  SFNVGFQNTGFNSGAGN 459

Db      442  SFNVGFQNTGFNSGAGN 459

RESULT 13
S70787
N:curlin nucleator protein csGB precursor - Escherichia coli (strain K-12)
X:Alternate names: csGB protein; curlin nucleation component; minor curlin protein
C:Species: Escherichia coli
C>Date: 12-Feb-1998 #sequence_revision 20-Feb-1998 #text_change 01-Mar-2002
C:Accession: S70787; F64846
R:Hammar, M.; Arngvist, A.; Bian, Z.; Olsen, A.; Normark, S.
Mol. Microbiol. 18, 661-670, 1995
A:Title: Expression of two csG operons is required for production of fibronectin- and Cor
A:Reference number: S70783; MUID:96414468; PMID:8817489
A:Accession: S70787
A:Status: nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-151 <HAM>
A:Cross-references: EMBL:X90754; NID:g1147558; PIDN:CAA62281.1; PID:g1147563
A:Experimental source: strain K12, substrain W3110
A:Note: The nucleotide sequence was submitted to the EMBL Data Library, August 1995
R:Blattner, F.R.; Plunkett III, G.; Bloch, C.A.; Perna, N.T.; Burland, V.; Riley, M.; Col
.A.; Rose, D.J.; Mau, B.; Shao, Y.
Science 277, 1453-1462, 1997
A:Title: The complete genome sequence of Escherichia coli K-12.
A:Reference number: A64720; MUID:97426617; PMID:9278503
A:Accession: F64846
A:Status: nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-151 <BLAT>
A:Cross-references: GB:AB000205; GB:U00096; NID:g1787265; PIDN:AAC74125.1; PID:g1787278;
A:Experimental source: strain K-12, substrain MG1655
C:Genetics:
A:Gene: csGB
A:Map position: 23.15
C:Function:
A:Description: minor component of wild-type curli; interaction between CsgA and CsgB tri
A:Note: curli are thin, coiled fibers expressed on the surface of Escherichia coli that i
and H-kininogen; in the absence of CsgA, CsgB can self-assemble into polymers
F:1-21/Domain: signal sequence #status predicted <SIG>
F:22-151/Product: minor curlin chain #status predicted <MAT>

Query Match 11.5%; Score 89.5; DB 2; Length 151;
Best Local Similarity 28.1%; Pred. No. 0.92;
Matches 32; Conservative 15; Mismatches 62; Indels 5; Gaps 2;

Qy      38  SSGPDSLTLSIYQGSANAALQSDARKSETTITQSGYGNAGDYDOLVTRVTHEMAHAF 97
      99  AAGVYLANSEYNF---AVNELSKSFNQAAIIGAGTWNNSAQLRGGSKLLAVAQEGS 76
      98  RNNATIDOWNAKNSDITVGYGNNALVNQTPASDSVMVRQVGFQNNATANYQ 151
      77  SNRAKIDQGTGYNL-AVIDQAGSANDASISQAGYGNNTAMIIQKSGSKNANITQY 129

Db      98  RNNATIDOWNAKNSDITVGYGNNALVNQTPASDSVMVRQVGFQNNATANYQ 151
      77  SNRAKIDQGTGYNL-AVIDQAGSANDASISQAGYGNNTAMIIQKSGSKNANITQY 129

RESULT 14
C90806
N:minor curlin subunit precursor CsgB [imported] - Escherichia coli (strain O157:H7, substi
C:Species: Escherichia coli
C>Date: 18-Jul-2001 #sequence_revision 18-Jul-2001 #text_change 18-Jul-2001
R:Hayashi, T.; Makino, K.; Ohnishi, M.; Kurokawa, K.; Ishii, K.; Yokoyama, K.; Han, C.G.;
gasawara, N.; Yasunaga, T.; Kuhara, S.; Shiba, T.; Hattori, M.; Shinagawa, H.
DNA Res. 8, 11-22, 2001
A:Title: Complete genome sequence of enterohemorrhagic Escherichia coli O157:H7 and genom

```

A:Reference number: A99629; MUID:21156231; PMID:11258796

A:Accession: C90806

A>Status: preliminary

A:Molecule type: DNA

A:Residues: 1-151 <HAY>

A:Cross-references: GB:BA000007; PIDN:BA034842.1; PID:gl3360879; GSPDB:GN00154

A:Experimental source: strain O157:H7, substrain RIMD 0509952

C:Genetics:

A:Gene: ECs1419

Query Match 11.5%; Score 89.5; DB 2; Length 151;

Best Local Similarity 28.1%; Pred. No. 0.92;

Matches 32; Conservative 15; Mismatches 62; Indels 5; Gaps 2;

QY 38 SSGPDSLSTIYQYGSANAALQSDARKSETTITQSGYGNAGADYDQLVTRVVTHEMAHAF 97

Db 21 AAGYDLANSEYNF-----AVNELSKSFNQAAIIGQAGTNNSAQLRQGGSKLLAVVAQEGS 76

QY 98 RNNATIDQWNAKNSDITVGQYGGNNAALVNOTASDSSVMVRQVGFGNNTATNOY 151

Db 77 SNRAKIDQTGDYNL-AVIDQAGSANDASISQAGYGNNTAMIIQKSGNKNANITQY 129

#### RESULT 15

G85665

curlin minor chain precursor, CsgA homolog [imported] - Escherichia coli (strain O157:H7

C:Species: Escherichia coli

C>Date: 16-Feb-2001 #sequence\_revision 16-Feb-2001 #text\_change 14-Sep-2001

C:Accession: G85665

R;Perna, N.T.; Plunkett III, G.; Burland, V.; Mau, B.; Glasner, J.D.; Rose, D.J.; Mayhew

Iller, L.; Grotbeck, E.J.; Davis, N.W.; Lim, A.; Dimalanta, E.; Potamousis, K.; Apodaca,

Nature 409, 529-533, 2001

A:Title: Genome sequence of enterohemorrhagic Escherichia coli O157:H7.

A:Reference number: A85480; MUID:21074935; PMID:11206551

A:Accession: G85665

A>Status: preliminary

A:Molecule type: DNA

A:Residues: 1-151 <STO>

A:Cross-references: GB:AE005174; NID:gl2514573; PIDN:AGS55787.1; GSPDB:GN00145; UNGP:Z16

A:Experimental source: strain O157:H7, substrain EDL933

C:Genetics:

A:Gene: csgB

Query Match 11.5%; Score 89.5; DB 2; Length 151;

Best Local Similarity 28.1%; Pred. No. 0.92;

Matches 32; Conservative 15; Mismatches 62; Indels 5; Gaps 2;

QY 38 SSGPDSLSTIYQYGSANAALQSDARKSETTITQSGYGNAGADYDQLVTRVVTHEMAHAF 97

Db 21 AAGYDLANSEYNF-----AVNELSKSFNQAAIIGQAGTNNSAQLRQGGSKLLAVVAQEGS 76

QY 98 RNNATIDQWNAKNSDITVGQYGGNNAALVNOTASDSSVMVRQVGFGNNTATNOY 151

Db 77 SNRAKIDQTGDYNL-AVIDQAGSANDASISQAGYGNNTAMIIQKSGNKNANITQY 129

Search completed: August 2, 2004, 14:56:25

Job time : 9.4 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: August 2, 2004, 14:36:12 ; Search time 5.3 Seconds  
(without alignments)  
1483.508 Million cell updates/sec

Title: US-09-543-407-28

Perfect score: 775

Sequence: 1 MKLLKVAFAAIVVSGSALA.....DSSVMVRQVGFNNATANQY 151

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 141681 seqs, 52070155 residues

Total number of hits satisfying chosen parameters: 141681

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt\_42.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	683	88.1	151	1	CSGA_SALTY
2	509	65.7	151	1	CSGA_ECOLI
3	487.5	62.9	151	1	CSGA_ECO57
4	94.5	12.2	151	1	CSGB_SALTY
5	91.5	11.8	602	1	GP63_LEIMA
6	89.5	11.5	151	1	CSGB_ECOLI
7	89.5	11.5	1322	1	ICFA_PANAN
8	89.5	11.5	1567	1	ICEN_XANCT
9	87.5	11.3	1034	1	ICEN_PANAN
10	87.5	11.3	1258	1	ICEN_BRWHE
11	86.5	11.2	151	1	CSGB_SALTI
12	86	11.1	590	1	GP63_LEIDO
13	84.5	10.9	646	1	GP63_LEIME
14	84	10.8	599	1	GP63_LEICH
15	84	10.8	1608	1	HLVA_SERMA
16	83.5	10.8	856	1	ALE7_AGOVI
17	83	10.7	1656	1	OMPE_RICJA
18	81.5	10.5	392	1	HME1_HUMAN
19	81.5	10.5	697	1	SIM_DROME
20	81.5	10.5	1028	1	OVO_DROME
21	80	10.3	331	1	OMB2_NEIME
22	80	10.3	760	1	YBIL_ECOLI
23	79.5	10.3	475	1	ABPB_STRKO
24	79.5	10.3	548	1	CEAK_ECOLI
25	79.5	10.3	1210	1	ICEN_PSEFL
26	79	10.2	678	1	YF48_MICTU
27	79	10.2	823	1	NSP1_PEA
28	78.5	10.1	232	1	DHN3_PEA
29	78.5	10.1	641	1	IMD_ARTGO
30	77.5	10.0	1063	1	SPT5_YEAST
31	77.5	10.0	1148	1	ICEV_PESX
32	77.5	10.0	1185	1	MAPX_DROME
33	77.5	10.0	1196	1	ICEV_PESX

34	77	9.9	204	1	CORA_MEDSA
35	77	9.9	376	1	P33_MYCPE
36	77	9.9	439	1	ACMA_LACLA
37	76.5	9.9	487	1	Y442_MYCTU
38	76.5	9.9	881	1	YFCU_ECOLI
39	76	9.8	147	1	HFAA_CAUCR
40	76	9.8	362	1	P35_MYCPE
41	76	9.8	542	1	SCWB_YEAST
42	76	9.8	1571	1	C3G_DROME
43	75.5	9.7	163	1	HCY_NATPH
44	75.5	9.7	394	1	OMSI_SALTI
45	75.5	9.7	424	1	COAA_BFPD

## ALIGNMENTS

RESULT 1					
ID	CSGA_SALTY	STANDARD;	PRT;	151	AA.
AC	P55225;				
DT	01-OCT-1996 (Rel. 34, Created)				
DT	01-OCT-1996 (Rel. 34, Last sequence update)				
DT	10-OCT-2003 (Rel. 42, Last annotation update)				
DE	Major curlin subunit, precursor (Fimbrin SBE17).				
GN	CSGA OR AGFA CR STM1144 OR STY1181 OR T1776.				
OS	Salmonella typhimurium,				
OS	Salmonella typhi, and				
OC	Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;				
OC	Enterobacteriaceae; Salmonella.				
OX	NCBI_TaxID=602, 601, 592;				
RP	SEQUENCE FROM N.A.				
RC	SPECIES=S.typhimurium; STRAIN=SR-11;				
RX	MEDLINE=98117058; PubMed=9457880;				
RA	Romling U., Bian Z., Hammar M., Sierralta W.D., Normark S.;				
RT	"Curli fibers are highly conserved between Salmonella typhimurium and				
RT	Escherichia coli with respect to operon structure and regulation.";				
RL	J. Bacteriol. 180:722-731(1998).				
RL	[2]				
RP	SEQUENCE FROM N.A.				
RC	SPECIES=S.typhimurium; STRAIN=LT2 / SGSC1412 / ATCC 700720;				
RX	MEDLINE=21534947; PubMed=11677609;				
RA	McClelland M., Sanderson K.E., Spieth J., Clifton S.W., Latreille P.,				
RA	Courtney L., Porwollik S., Ali J., Dente M., Du F., Hou S., Layman D.,				
RA	Leonard S., Nguyen C., Scott K., Holmes A., Grewal N., Mulvaney E.,				
RA	Ryan E., Sun H., Florea L., Miller W., Stoneking T., Nhan W.,				
RA	Waterston R., Wilson R.K.;				
RT	"Complete genome sequence of Salmonella enterica serovar Typhimurium				
RT	LT2.";				
RL	Nature 413:852-856(2001).				
RL	[3]				
RP	SEQUENCE FROM N.A.				
RC	SPECIES=S.typhi; STRAIN=CT18;				
RX	MEDLINE=21534947; PubMed=11677608;				
RA	Parkhill J., Dougan G., James K.D., Holden M.T.G., Sebahia M.,				
RA	Churcher C., Mangall K.L., Bentley S.D., Chillingworth T., Connor P.,				
RA	Baker S., Basham D., Brooks K., Chillingworth T., White N., Farrar J.,				
RA	Cronin A., Davis P., Davies R.M., Dowd L., White N., Farrar J.,				
RA	Feltwell T., Hamlin N., Haque A., Hien T.T., Holroyd S., Jagels K.,				
RA	Krogh A., Larsen T.S., Leather S., Moulé S., O'Gaora P., Parry C.,				
RA	Quail M.A., Rutherford K., Simmonds M., Skelton J., Stevens K.,				
RT	"Complete genome sequence of a multiple drug resistant Salmonella				
RT	enterica serovar Typhi CT18.";				
RL	Nature 413:848-852(2001).				
RL	[4]				
RP	SEQUENCE FROM N.A.				
RC	SPECIES=S.typhi; STRAIN=Ty2 / ATCC 700931;				
RX	MEDLINE=22531367; PubMed=12644504;				
RA	Deng W., Liou S.-R., Plunkett G. III, Mayhew G.F., Rose D.J.,				
RA	Burland V., Kodoyanni V., Schwartz D.C., Blattner F.R.;				

Q07202	medicago sa
Q50368	mycoplasma
Q9cit4	lactococcus
P42611	mycobacteri
P77196	escherichia
P27342	caulobacter
Q50367	mycoplasma
P53189	saccharomyc
Q77086	drosophila
P39442	natronomona
Q56110	salmonella
P03661	bacterioph

RT "Comparative genomics of Salmonella enterica serovar Typhi strains Ty2  
 RL and CT18.";  
 RN J. Bacteriol. 185:2330-2337(2003).  
 [5]

RP SEQUENCE FROM N.A.  
 RC SPECIES=enteritidis; STRAIN=27655-3B;  
 RX MEDLINE=96146512; PubMed=8550497;  
 RA Collinson S.K., Clouthier S.C., Doran J.L., Baner P.A., Kay W.W.;  
 RT "Salmonella enteritidis agfBAC operon encoding thin, aggregative  
 RL fimbriae.";  
 RN J. Bacteriol. 178:662-667(1996).  
 [6]

RP SEQUENCE OF 21-151 FROM N.A.  
 RC SPECIES=enteritidis; STRAIN=27655-3B;  
 RX MEDLINE=91310596; PubMed=1677357;  
 RA Collinson S.K., Emeody L., Mueller K.-M., Trust T.J., Kay W.W.;  
 RT "Purification and characterization of thin, aggregative fimbriae from  
 RL Salmonella enteritidis.";  
 RN J. Bacteriol. 173:4773-4781(1991).  
 [7]

RP SEQUENCE OF 21-33.  
 RC SPECIES=enteritidis; STRAIN=27655-3B;  
 RX MEDLINE=91310596; PubMed=1677357;  
 RA Collinson S.K., Emeody L., Mueller K.-M., Trust T.J., Kay W.W.;  
 RT "Purification and characterization of thin, aggregative fimbriae from  
 RL Salmonella enteritidis.";  
 RN J. Bacteriol. 173:4773-4781(1991).  
 [7]

CC -!- FUNCTION: CURLIN IS THE STRUCTURAL SUBUNIT OF THE CURLI. CURLI ARE  
 COILED SURFACE STRUCTURES THAT ASSEMBLE PREFERENTIALLY AT GROWTH  
 TEMPERATURES BELOW 37 DEGREES CELSIUS. CURLI CAN BIND TO  
 FIBRONECTIN.

CC -!- SIMILARITY: BELONGS TO THE CSGA/CSGB FAMILY.

CC This SWISS-PROT entry is copyrighted. It is produced through a collaboration  
 between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 the European Bioinformatics Institute. There are no restrictions on its  
 use by non-profit institutions as long as its content is in no way  
 modified and this statement is not removed. Usage by and for commercial  
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC -----  
 CC EMBL: AJ002301; CAA05317.1; -  
 CC EMBL: AB008749; AAL20074.1; -  
 CC EMBL: AL627269; CAD08268.1; -  
 CC EMBL: AB016840; AAC06939.1; -  
 CC EMBL: U43280; AAC43599.1; -  
 CC FIR: JC6039; JC6039.  
 CC StyGene; SGI0608; csGA.  
 CC FimBria; Signal; Complete proteome.  
 KW SIGNAL 1 20  
 FT CHAIN 21 151  
 FT CONFLICT 134 151  
 FT REF. 6)  
 FT MAJOR CURLIN SUBUNIT.  
 FT SYMVQVGGNNATANYQ -> DSYTQVAS (IN  
 FT REF. 6)  
 FT SEQUENCE 151 AA; 15305 MW; 87DAC0D16B621359 CRC64;

Query Match 88.1%; Score 683; DB 1; Length 151;  
 Best Local Similarity 90.1%; Pred. No. 1.3e-51;  
 Matches 136; Conservative 1; Mismatches 14; Indels 0; Gaps 0;

QY 1 MKLLKVAFAAIVVSGSALAGVVPQGGNGGNGSGPDSLTICYGSANALALQ 60

DB 1 MKLLKVAFAAIVVSGSALAGVVPQGGNGGNGSGPDSLTICYGSANALALQ 60

QY 61 SDARKSETTITGSGYNGADYDQLVTRVVTVMHAFRNATIDQWNAKNSDITVQYGG 120

DB 61 SDARKSETTITGSGYNGADYDQLVTRVVTVMHAFRNATIDQWNAKNSDITVQYGG 120

QY 121 NNAALVNOTASDSSVMVQVGGNNATANYQ 151

DB 121 NNAALVNOTASDSSVMVQVGGNNATANYQ 151

RESULT 2  
 ID CSGA\_ECOLI STANDARD; PRT; 151 AA.  
 AC P28307;  
 DT 01-DEC-1992 (Rel. 24, Created)  
 DT 01-OCT-1996 (Rel. 34, Last sequence update)  
 DT 28-FEB-2003 (Rel. 41, Last annotation update)  
 DE Major curlin subunit precursor.  
 GN CSGA OR B1042.  
 OS Escherichia coli.  
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;  
 OC Enterobacteriaceae; Escherichia.  
 OX NCBI\_TaxID=562;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=K12 / W3110;  
 RX MEDLINE=93211294; PubMed=8459772;  
 RA Olsen A., Arngvist A.;  
 RT "The RpoS sigma factor relieves H-NS-mediated transcriptional  
 RL repression of csGA, the subunit gene of fibronectin-binding curli in  
 RL Escherichia coli.";  
 RL Mol. Microbiol. 7:523-536(1993).  
 [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=K12 / MC4100;  
 RX MEDLINE=96414468; PubMed=8817489;  
 RA Hammar M., Arngvist A., Bian Z., Olsen A., Normark S.;  
 RT "Expression of two csG operons is required for production of  
 RL fibronectin- and congo red-binding curli polymers in Escherichia coli  
 RL K-12.";  
 RL Mol. Microbiol. 18:661-670(1995).  
 [3]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=K12 / MCG1655;  
 RX MEDLINE=97426617; PubMed=9278503;  
 RA Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,  
 RA Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,  
 RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,  
 RA Mau B., Shao Y.;  
 RT "The complete genome sequence of Escherichia coli K-12.";  
 RL Science 277:1453-1474(1997).  
 [4]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=K12;  
 RX MEDLINE=97061202; PubMed=8905232;  
 RA Oshima T., Aiba H., Baba T., Fujita K., Hayashi K., Honjo A.,  
 RA Ikemoto K., Inada T., Itoh T., Kajihara M., Kanai K., Kashimoto K.,  
 RA Kimura S., Kitagawa M., Makino K., Masuda S., Miki T., Mizobuchi K.,  
 RA Mori H., Motomura K., Nakamura Y., Nishimoto H., Nishio Y., Saito N.,  
 RA Sampei G., Seki Y., Tagami H., Takemoto K., Wada C., Yamamoto Y.,  
 RA Yano M., Horiuchi T.;  
 RT "A 718-kb DNA sequence of the Escherichia coli K-12 genome  
 RL corresponding to the 12.7-28.0 min region on the linkage map.";  
 RL DNA Res. 3:137-155(1996).  
 [5]  
 RP SEQUENCE OF 21-40.  
 RC STRAIN=K12 / YMEI;  
 RX MEDLINE=93023873; PubMed=1357528;  
 RA Arngvist A., Olsen A., Pfeifer J., Russell D.G., Normark S.;  
 RT "The Crl protein activates cryptic genes for curli formation and  
 RL fibronectin binding in Escherichia coli HB101.";  
 RL Mol. Microbiol. 6:2443-2452(1992).  
 [6]  
 RP SEQUENCE OF 21-31.  
 RC MEDLINE=91310586; PubMed=1677357;  
 RA Collinson S.K., Emeody L., Trust T.J., Kay W.W.;  
 RT "Purification and characterization of thin, aggregative fimbriae from  
 RL Salmonella enteritidis.";  
 RL J. Bacteriol. 173:4773-4781(1991).  
 CC -!- FUNCTION: CURLIN IS THE STRUCTURAL SUBUNIT OF THE CURLI. CURLI ARE  
 COILED SURFACE STRUCTURES THAT ASSEMBLE PREFERENTIALLY AT GROWTH  
 TEMPERATURES BELOW 37 DEGREES CELSIUS. CURLI CAN BIND TO  
 FIBRONECTIN.



RL J. Bacteriol. 180:722-731(1998).  
RN [2]  
RC SEQUENCE FROM N.A.  
RX MEDLINE=21534948; PubMed=11677609;  
RA McClelland M., Sanderson K.E., Spieth J., Clifton S.W., Latreille P.,  
RA Courtney L., Forwell S., Ali J., Dante M., Du F., Hou S., Layman D.,  
RA Leonard S., Nguyen C., Scott K., Holmes A., Grewal N., Mulvaney E.,  
RA Ryan E., Sun H., Florea L., Miller W., Stoneking T., Nhan M.,  
RA Waterston R., Wilson R.K.;  
RT "Complete genome sequence of Salmonella enterica serovar Typhimurium  
LT2";  
RL Nature 413:852-856(2001).  
RN [3]  
RC SEQUENCE FROM N.A.  
RX SPECIES=Salmonella; STRAIN=27655-3B;  
RX MEDLINE=96146512; PubMed=8550497;  
RA Collinson S.K., Clouthier S.C., Doran J.L., Baner P.A., Kay W.W.;  
RT "Salmonella enteritidis agfBAC operon encoding thin, aggregative  
fimbriae";  
RL J. Bacteriol. 178:662-667(1996).  
RN [4]  
RC COILED SURFACE STRUCTURES THAT ASSEMBLE PREFERENTIALLY AT GROWTH  
TEMPERATURES BELOW 37 DEGREES CELSIUS. CURLI CAN BIND TO  
FIBRONECTIN. THE MINOR SUBUNIT IS THE NUCLEATION COMPONENT OF  
CURLIN MONOMERS.  
CC  
CC -!- SIMILARITY: BELONGS TO THE CSGA/CSGB FAMILY.  
CC  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
between the Swiss Institute of Bioinformatics and the EMBL outstation -  
the European Bioinformatics Institute. There are no restrictions on its  
use by non-profit institutions as long as its content is in no way  
modified and this statement is not removed. Usage by and for commercial  
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC  
CC EMBL; AJ002301; CAA05316.1; -  
CC EMBL; AE008749; AAL20073.1; -  
CC EMBL; U43280; AAC43598.1; -  
CC FIRM; JC6040; JC6040.  
CC Stygenia; Signal; Complete proteome.  
KW Fimbria; Signal; 21 POTENTIAL.  
FT SIGNAL 1 21  
FT CHAIN 22 151 MINOR CURLIN SUBUNIT.  
SQ SEQUENCE 151 AA; 16182 MW; C0FC5430E6DD361D CRC64;  
  
Query Match 12.2%; Score 94.5; DB 1; Length 151;  
Best Local Similarity 26.3%; Pred.No.0.19; Indels 37; Gaps 4;  
Matches 35; Conservative 17; Mismatches 41;  
  
QY 10 AAIWVGSSALAGVVPQWGGGNGGNGSGPDTLSIYQVSANAALQSDARKSETT 69  
DB 58 ARVRQEGSKLLSVISO--EGGNRAKVQDAGNYNFAIETGTGNAN-----DAS 103  
QY 70 ITQSGYNGADYDQLYTRVTVTHEVAFRNATIDQWNAKNSDITVGYGGNNAALVNOT 129  
DB 104 ISQSAYGNSA-----AIIQKSGNKNANIT--QIGTKTAVVVK 140  
  
QY 130 ASDSSVMVRQ 139  
DB 141 QSHWAIKRTQ 150  
  
RESULT 5  
GP63 LEIMA  
ID GP63 LEIMA STANDARD; PRT; 602 AA.  
AC P08148; P15906;  
DT 01-AUG-1988 (Rel. 08, Created)  
DT 01-APR-1990 (Rel. 14, Last sequence update)  
DT 10-OCT-2003 (Rel. 42, Last annotation update)  
DE Leishmanolysin precursor (EC 3.4.24.36) (Cell surface protease)  
DE (Major surface glycoprotein) (GP63 protein) (promastigote surface  
DE endopeptidase).

GN  
OS Leishmania major.  
OC Eukaryota; Euglenozoa; Kinetoplastida; Trypanosomatidae; Leishmania.  
OX NCBI\_TaxID=5664;  
RN [1]  
RP SEQUENCE FROM N.A., AND SEQUENCE OF 101-123.  
RX MEDLINE=88154764; PubMed=3346625;  
RA Button L.L., McMaster W.R.;  
RT "Molecular cloning of the major surface antigen of Leishmania.";  
RL J. Exp. Med. 167:724-729(1988).  
RN [2]  
RP REVISIONS.  
RA Button L.L., McMaster W.R.;  
RL J. Exp. Med. 171:589-589(1990).  
RN [3]  
RP GPI-ANCHOR.  
RX MEDLINE=91009116; PubMed=2145267;  
RX Schneider P., Ferguson M.A.J., McConville M.J., Mehler A.,  
RX Homans S.W., Bordier C.;  
RT "Structure of the glycosyl-phosphatidylinositol membrane anchor of  
the Leishmania major promastigote surface protease.";  
RL J. Biol. Chem. 265:16955-16964(1990).  
RN [4]  
RP X-RAY CRYSTALLOGRAPHY (2.6 ANGSTROMS).  
RX MEDLINE=95406217; PubMed=7675788;  
RX Schlagenhauf E., Eiges R., Metcalf P.;  
RT "Crystallization and preliminary X-ray diffraction studies of  
leishmanolysin, the major surface metalloproteinase from Leishmania  
major";  
RL proteins 22:58-66(1995).  
RN [5]  
RP X-RAY CRYSTALLOGRAPHY (1.86 ANGSTROMS).  
RX MEDLINE=98416698; PubMed=9739094;  
RX Schlagenhauf E., Eiges R., Metcalf P.;  
RT "The crystal structure of the Leishmania major surface proteinase  
leishmanolysin";  
RL Structure 6:1035-1046(1998).  
CC -!- FUNCTION: Has an integral role during the infection of macrophages  
in the mammalian host.  
CC -!- CATALYTIC ACTIVITY: Preference for hydrophobic residues at P1 and  
P1' and basic residues at P2 and P3'. A model nonapeptide is  
cleaved at -Ala-Tyr-|-Leu-Lys-Lys-.  
CC -!- COFACTOR: Binds 1 zinc ion per subunit.  
CC -!- SUBCELLULAR LOCATION: Attached to the membrane by a GPI-anchor.  
CC -!- PTM: THE PHOSPHATIDYLINOSITOL MOIETY OF THE GPI-ANCHOR CONTAINS A  
FULLY SATURATED, UNBRANCHED 1-O-ALKYL CHAIN (MAINLY C24:0) AND A  
MIXTURE OF FULLY SATURATED UNBRANCHED 2-O-ACYL CHAINS (C12:0,  
C14:0, C16:0, AND C18:0).  
CC -!- SIMILARITY: Belongs to peptidase family M8.  
CC  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
between the Swiss Institute of Bioinformatics and the EMBL outstation -  
the European Bioinformatics Institute. There are no restrictions on its  
use by non-profit institutions as long as its content is in no way  
modified and this statement is not removed. Usage by and for commercial  
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC  
CC EMBL; Y00647; CAA68673.1; -  
CC FIRM; PLO221; PLO221.  
CC PDB; 1LML; 17-SEP-97.  
CC MEROPS; M08.001; -  
CC InterPro; IPR006025; Pept M Zn BS.  
CC InterPro; IPR001577; Peptidase M8.  
CC Pfam; PF01457; Peptidase M8; 1.  
CC PRINIS; PR00782; LSHMANOLYSIN.  
CC PROSITE; PS00142; ZINC\_PROTEASE; 1.  
CC Hydrolase; Metalloprotease; Glycoprotein; Metal-binding; Zinc;  
KW Zymogen; Signal; Cell adhesion; GPI-anchor; 3D-structure; Lipoprotein.  
FT SIGNAL 1 39 POTENTIAL.  
FT PROPEP 40 100 ACTIVATION PEPTIDE.  
FT CHAIN 101 577 LEISHMANOLYSIN.  
FT PROPEP 578 602 REMOVED IN MATURE FORM.

FT	STRAND	421	425
FT	STRAND	428	429
FT	HELIX	435	437
TURN		443	444
FT	STRAND	445	446
FT	TURN	450	454
FT	STRAND	458	465
FT	TURN	466	467
FT	HELIX	470	472
FT	TURN	475	477
FT	HELIX	478	480
FT	TURN	485	486
FT	STRAND	487	494
FT	STRAND	496	496
FT	STRAND	506	516
FT	TURN	517	520
FT	STRAND	521	525
FT	TURN	527	528
FT	STRAND	533	534
FT	TURN	537	538
FT	STRAND	540	542
FT	HELIX	543	545
FT	TURN	546	546
FT	STRAND	550	550
FT	TURN	552	553
FT	STRAND	555	557
FT	HELIX	561	565
FT	TURN	566	567
FT	HELIX	569	572
FT	TURN	573	573
Q	SEQUENCE	602 AA;	982EF3245D87C43E CRC64;

Query Match 11.8%; Score 91.5; DB 1; Length 602;  
Best Local Similarity 41.7%; Pred. No. 1.6;  
Matches 25; Conservative

Qy	81	YDQLVTRVTHVTHMAHAF-----RNNATIDQWNAKSDITV-----GGYG 119
Db	254	YDQLVTRVTHVTHMAHALGFGSPFFEDARIVANVENVRGKFDVPEVINSSTAVAKAREQYG 313

RESULT 6

CSGB	ECOLI	STANDARD;	PRT;	151 AA.
ID	CSGB_ECOLI	STANDARD;	PRT;	151 AA.
AC	P39828;			
DT	01-FEB-1995 (Rel. 31, Created)			
DD	01-OCT-1996 (Rel. 34, Last sequence update)			
DE	28-FEB-2003 (Rel. 41, Last annotation update)			
DE	Minor curlin subunit precursor			
GN	CSGB OR B1041 OR Z1675 OR ECS1419.			
OS	Escherichia coli, and			
OS	Escherichia coli O157:H7.			
OC	Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;			
OC	Enterobacteriaceae; Escherichia.			
OX	NCBI_TaxID=562, 83334;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN=K12 / MC4100;			
RC	MEDLINE=96414469; PubMed=8817489;			
RC	MEDLINE=97426617; PubMed=9278503;			
RA	Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,			
RA	Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,			
RA	Gregor J., Davis N.W., Kirkpatrick H.A., Goeden W.A., Rose D.J.,			
RA	Mau B., Shao Y.,			
RT	"The complete genome sequence of Escherichia coli K-12."			
RL	Mol. Microbiol. 18:661-670(1995).			
RN	[2]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN=K12 / MG1655;			
RC	MEDLINE=97426617; PubMed=9278503;			
RA	Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,			
RA	Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,			
RA	Gregor J., Davis N.W., Kirkpatrick H.A., Goeden W.A., Rose D.J.,			
RA	Mau B., Shao Y.,			
RT	"The complete genome sequence of Escherichia coli K-12."			

```

RL  Science 277:1453-1474 (1997).
RN  [3]
RP  SEQUENCE FROM N.A.
RX  STRAIN=K12;
RX  MEDLINE=97061202; PubMed=8905232;
RA  Oshima T., Aiba H., Baba T., Fujita K., Hayashi K., Honjo A.,
RA  Ikemoto K., Inada T., Itoh T., Kajihara M., Kanai K., Kashimoto K.,
RA  Kimura S., Kitagawa M., Makino K., Masuda S., Miki T., Mizobuchi K.,
RA  Mori H., Motomura K., Nakamura Y., Nishio Y., Saico N.,
RA  Sampei G., Seki Y., Tagami H., Takemoto K., Wada C., Yamamoto Y.,
RA  Yano M., Horiuchi T.;
RT  "A 718-Kb DNA sequence of the Escherichia coli K-12 genome
RT  corresponding to the 12.7-28.0 min region on the linkage map.";
RL  DNA Res. 3:137-155 (1996).
RN  [4]
RP  SEQUENCE FROM N.A.
RX  STRAIN=O157:H7 / EDL933 / ATCC 700927;
RX  MEDLINE=21074935; PubMed=11206551;
RA  Perna N.T., Plunkett G. III, Burland V., Mau B., Glasner J.D.,
RA  Rose D.J., Mayhew G.F., Evans P.S., Gregor J., Kirkpatrick H.A.,
RA  Posfai G., Hackett J., Klink S., Boutin A., Shao Y., Miller L.,
RA  Grobeck E.J., Davis N.W., Lim A., Dimalanta E.T., Potamousis K.,
RA  Apodaca J., Anantharaman T.S., Lin J., Yen G., Schwartz D.C.,
RA  Welch R.A., Blattner F.R.;
RT  "Genome sequence of enterohaemorrhagic Escherichia coli O157:H7.";
RL  Nature 409:529-533 (2001).
RN  [5]
RP  SEQUENCE FROM N.A.
RX  STRAIN=O157:H7 / RIMD 0509952;
RX  MEDLINE=21156231; PubMed=11258796;
RA  Hayashi T., Makino K., Onishi M., Kurokawa K., Ishii K., Yokoyama K.,
RA  Han C.-G., Ohtsubo E., Nakayama K., Murata T., Tanaka M., Tobe T.,
RA  Tada T., Takami H., Honda T., Sasakawa C., Ogasawara N., Yasunaga T.,
RA  Kuhara S., Shiba T., Hattori M., Shinagawa H.;
RT  "Complete genome sequence of enterohemorrhagic Escherichia coli
RT  O157:H7 and genomic comparison with a laboratory strain K-12.";
RL  DNA Res. 8:11-22 (2001).
RN  [6]
RP  SEQUENCE OF 1-21 FROM N.A.
RX  STRAIN=K12;
RX  MEDLINE=95157246; PubMed=7854117;
RA  Arngvist A., Olsen A., Normark S.;
RT  "Sigma S-dependent growth-phase induction of the csgBA promoter in
RT  Escherichia coli can be achieved in vivo by sigma 70 in the absence
RT  of the nucleoid-associated protein H-NS.";
RL  Mol. Microbiol. 13:1021-1032 (1994).
CC  -!- FUNCTION: CURLIN IS THE STRUCTURAL SUBUNIT OF THE CURLI. CURLI ARE
CC  COILED SURFACE STRUCTURES THAT ASSEMBLE PREFERENTIALLY AT GROWTH
CC  TEMPERATURES BELOW 37 DEGREES CELSIUS. CURLI CAN BIND TO
CC  FIBRONECTIN. THE MINOR SUBUNIT IS THE NUCLEATION COMPONENT OF
CC  CURLIN MONOMERS.
CC  -!- SIMILARITY: BELONGS TO THE CSGA/CSGB FAMILY.
CC  -----
CC  This SWISS-PROT entry is copyright. It is produced through a collaboration
CC  between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC  the European Bioinformatics Institute. There are no restrictions on its
CC  use by non-profit institutions as long as its content is in no way
CC  modified and this statement is not removed. Usage by and for commercial
CC  entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC  or send an email to license@isb-sib.ch).
CC  -----
CC  EMBL; X90754; CAA62281.1; -
CC  DR  EMBL; AE000205; AAC74125.1; -
CC  DR  EMBL; D90741; BAA35831.1; -
CC  DR  EMBL; AE005315; AGS55787.1; -
CC  DR  EMBL; AP002554; BAB34842.1; -
CC  DR  PIR; C90806; C90806.
CC  DR  PIR; G85665; G85665.
CC  DR  PIR; S70787; S70787.
CC  DR  EcoGene; EG12621; csgB.
CC  KW  Fimbria; Signal; Complete proteome.
CC  FT  SIGNAL 1 21 POTENTIAL.
CC  CHAIN 22 151 MINOR CURLIN SUBUNIT.

```

---

```

SQ  SEQUENCE 151 AA; 15882 MW; B18D266B964014B8 CRC64;
Query Match 11.5%; Score 89.5; DB 1; Length 151;
Best Local Similarity 28.1%; Pred. No. 0.5;
Matches 32; Conservative 15; Mismatches 62; Indels 5; Gaps 2;
Qy 38 SSGPSTLSIYQYGSNAALALQSDARKSETTITQSYGNGADYDQIVTRVTHENAHP 97
Db 21 AAGYDLANSYNF---AVNELSKSFNQAAIIGQAGTNNASQLRQSGSKLLAVQAQGS 76
Qy 96 RNNATIDOWNAKNSDITVGYGNGNAALVNOTASDSSVMVRQVGFNGNATANCY 151
Db 77 SNRAKIDQTDYNL-AVIDAGSANDASISQAGYGNATMIQKSGNKANITQY 129
RESULT 7
ID ICEA PANAN STANDARD; PRT; 1322 AA.
AC P20469;
DT 01-FEB-1991 (Rel. 17, Created)
DT 01-FEB-1991 (Rel. 17, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Ice nucleation protein inaa.
GN INAA.
OS Pantoea ananas (Erwinia uredovora).
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Pantoea.
OC NCBI_TaxID=553;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=90092494; PubMed=2599095;
RA Abe K., Watabe S., Emori Y., Watanabe M., Arai S.;
RT "An ice nucleation active gene of Erwinia ananas. Sequence similarity
RT to those of Pseudomonas species and regions required for ice
RT nucleation activity.";
RL FEBS Lett. 258:297-300 (1989).
CC -!- FUNCTION: Ice nucleation proteins enable bacteria to nucleate
CC crystallization in supercooled water.
CC -!- SUBCELLULAR LOCATION: Outer membrane (By similarity).
CC -!- DOMAIN: CONTAINS MANY IMPERFECT REPEATS OF THE CONSENSUS
CC OCTAPEPTIDE A-G-Y-G-S-T-X-T, FURTHER ON A 16-RESIDUE AND A
CC REGIONAL 48-RESIDUE PERIODICITY IS SUPERIMPOSED.
CC -!- SIMILARITY: Belongs to the bacterial ice nucleation protein
CC family.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; X17216; CAA35194.1; -
CC DR  PIR; S07053; S07053.
CC DR  HSSP; P06620; 11NA.
CC DR  InterPro; IPR000258; Ice_nucleatn.
CC DR  Pfam; PF00818; Ice_nucleation; 69.
CC DR  PRINTS; PR00327; ICENUCLEATN.
CC DR  PROSITE; PS00314; ICE_NUCLEATION; 49.
CC DR  Ice_nucleation; Repeat; Outer membrane.
CC DR  DOMAIN 162 1281; OCTAPEPTIDE PERIODICITY.
CC DR  SEQUENCE 1322 AA; 131094 MW; 89B0EE24AA837039 CRC64;
Query Match 11.5%; Score 89.5; DB 1; Length 1322;
Best Local Similarity 26.8%; Pred. No. 5.9;
Matches 34; Conservative 18; Mismatches 26; Indels 49; Gaps 7;
Qy 34 GGNSSGPDSTLSIYQYGSNAALALQSDARKSETTITQSYGNGADYDQIVTRVTHEM 93
Db 933 GSTTAGPDSLSL-TAGYGSTQTA-----GYNLSLTAGYGS----- 966
Qy 94 AHAFRNATIDQWNAKNSDITVGYG-----GNNAALV-----NQTASDSSVMVRQVGF 143

```



```
Db 967 -----TGTGQNSDLTTG-YGSTAGYESSLIAGYGSTGTATSPKSTIM--AGYG 1013
QY 144 NNATANO 150
Db 1014 SSGTARE 1020

RESULT 8
ICEN_XANCT
ID ICEN_XANCT STANDARD; PRT; 1567 AA.
AC P18127;
DT 01-NOV-1990 (Rel. 16, Created)
DT 01-NOV-1990 (Rel. 16, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Ice nucleation protein.
GN INAX.
OS Xanthomonas campestris (pv. translucens).
OC Bacteria; Proteobacteria; Gammaproteobacteria; Xanthomonadales;
OC Xanthomonadaceae; Xanthomonas.
OX NCBI_TaxID=343;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=X568;
RX MEDLINE=91080859; PubMed=2259339;
RA Zhao J., Orser C.S.;
RT "Conserved repetition in the ice nucleation gene inax from
RT Xanthomonas campestris pv. translucens."
RL Mol. Genet. 223:163-166(1990).
CC -!- FUNCTION: Ice nucleation proteins enable bacteria to nucleate
CC crystallization in supercooled water.
CC -!- SUBCELLULAR LOCATION: Outer membrane (By similarity).
CC -!- DOMAIN: CONTAINS 153 IMPERFECT REPEATS OF THE CONSENSUS
CC OCTAPEPTIDE A-G-Y-G-S-T-L-I; FURTHER ON A 16-RESIDUE AND A
CC REGIONAL 48-RESIDUE PERIODICITY IS SUPERIMPOSED.
CC -!- MISCELLANEOUS: A STRUCTURAL MODEL IS SUGGESTED IN WHICH THE ICE
CC NUCLEATION PROTEIN DISPLAYS A SYMMETRY RELATED TO THAT OF ICE.
CC -!- SIMILARITY: Belongs to the bacterial ice nucleation protein
CC family.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; X52970; CAA37140.1; -
CC HSSP; P06620; 11NA.
CC InterPro; IPR000258; Ice_nucleatn.
CC Pfam; PF00818; Ice_nucleation; 81.
CC PRINTS; PR00327; ICENUCLEATN.
CC PROSITE; PS00314; ICE_NUCLEATION; 57.
KW Ice nucleation; Repeat; Outer membrane.
SQ SEQUENCE 1567 AA; 152548 MW; C9B451D959ECAD63 CRC64;

Query Match 11.5%; Score 89.5; DB 1; Length 1567;
Best Local Similarity 26.4%; Pred. No. 7.2;
Matches 47; Conservative 23; Mismatches 59; Indels 49; Gaps 12;

QY 14 VSGSALAG-----VVPQWG--GGNH-----GGNSSGPDSTLSIYQGSANAAL 59
Db 205 YGSLTGTADQSLVAGYGSTAGDHSLLIAGYGSTGTAGSDSSI-LAGYGSTQTAAGR 263
QY 60 Q-----SDARKSETTITGSGY-----NGAD-----YDQLVTRVVTHEMAHAFFNNA 102
Db 264 STLTAGYGTQTQAQSRLLTSGYSTATSGSDSAVISGYGSTQTAGSESSLTGYGSTQT 323

103 IDQWAKNSDITVQYG-----GNNAALV-----NOTASDSSVMVQVGFNNATANO 150
324 A-----RKGSITAG-YGSTGTAGSDSALLIAGYGSTQTAGSESSLT--AGYGSTQTARK 374
```

```
RESULT 9
ICEN_PANAN
ID ICEN_PANAN STANDARD; PRT; 1034 AA.
AC Q47879;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Ice nucleation protein inau.
GN INAU.
OS Pantoea ananas (Erwinia uredovora).
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Pantoea.
OX NCBI_TaxID=553;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=KUIN-3;
RX MEDLINE=94264407; PubMed=7764866;
RA Michigami Y., Watabe S., Abe K., Obata H., Arai S.;
RT "Cloning and sequencing of an ice nucleation active gene of Erwinia
RT uredovora."
RL Biosci. Biotechnol. Biochem. 58:762-764(1994).
CC -!- FUNCTION: Ice nucleation proteins enable bacteria to nucleate
CC crystallization in supercooled water.
CC -!- SUBCELLULAR LOCATION: Outer membrane.
CC -!- DOMAIN: CONTAINS IMPERFECT REPEATS OF A CONSENSUS OCTAPEPTIDE
CC A-G-Y-G-S-T-X-T; FURTHER ON A 16-RESIDUE AND A REGIONAL 48-RESIDUE
CC PERIODICITY IS SUPERIMPOSED.
CC -!- MISCELLANEOUS: A STRUCTURAL MODEL IS SUGGESTED IN WHICH THE ICE
CC NUCLEATION PROTEIN DISPLAYS A SYMMETRY RELATED TO THAT OF ICE.
CC -!- SIMILARITY: Belongs to the bacterial ice nucleation protein
CC family.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; D14992; BAA03636.1; -
CC PIR; JC2143; JC2143.
CC HSSP; P06620; 11NA.
CC InterPro; IPR000258; Ice_nucleatn.
CC Pfam; PF00818; Ice_nucleation; 51.
CC PRINTS; PR00327; ICENUCLEATN.
CC PROSITE; PS00314; ICE_NUCLEATION; 34.
KW Ice nucleation; Repeat; Outer membrane.
FT DOMAIN 162 993 OCTAPEPTIDE PERIODICITY.
SQ SEQUENCE 1034 AA; 103378 MW; FA222523D333EADD CRC64;

Query Match 11.3%; Score 87.5; DB 1; Length 1034;
Best Local Similarity 25.2%; Pred. No. 6.6;
Matches 41; Conservative 18; Mismatches 41; Indels 63; Gaps 9;

QY 12 IVVGSALAGVVPQW--GGGNHNGGNSS-----GPDSTLSIYQGSANAAL 57
Db 161 IATYGSTLSGTHQSLIAGYGSTGTAGSDSTLIAGYGSTGTAGSDSTL-VAGYGSTQTA- 218
QY 58 ALQSDARKSETTITGSGYNGADYDQLVTRVVTHEMAHAFFNNAIDQWAKNSDITVQ 117
Db 219 -----GESSQVAGYGS-----TGTGKSGDLTAG- 243

118 YG-----GNNAALV-----NOTASDSSVMVQVGFNNATANO 150
244 YGSLTGTAGDSSLLIAGYGSTQTAGEDSSLT--AGYGSTQTAAK 284

RESULT 10
ICEN_ERWHE
ID ICEN_ERWHE STANDARD; PRT; 1258 AA.
AC P16239;
```

DT 01-APR-1990 (Rel. 14, Created)  
DT 01-APR-1990 (Rel. 14, Last sequence update)  
DT 16-OCT-2001 (Rel. 40, Last annotation update)  
DE Ice nucleation protein.  
GN ICEE.  
OS Erwinia herbicola.  
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;  
OC Enterobacteriaceae; Pantoea.  
OX NCBI\_TaxID=549;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=M1;  
RX MEDLINE=90152370; PubMed=2515997;  
RA Warren G.J., Corotto L.V.;  
RT "The consensus sequence of ice nucleation proteins from Erwinia  
herbicola, Pseudomonas fluorescens and Pseudomonas syringae.";  
RL Gene 85:239-242(1999).  
CC -!- FUNCTION: Ice nucleation proteins enable bacteria to nucleate  
crystallization in supercooled water.  
CC -!- SUBCELLULAR LOCATION: Outer membrane.  
CC -!- DOMAIN: CONTAINS 126 IMPERFECT REPEATS OF A CONSENSUS OCTAPEPTIDE  
A-G-Y-G-S-T-X-T. FURTHER ON A 16-RESIDUE AND A REGIONAL 48-RESIDUE  
PERIODICITY IS SUPERIMPOSED.  
CC -!- MISCELLANEOUS: A STRUCTURAL MODEL IS SUGGESTED IN WHICH THE ICE  
NUCLEATION PROTEIN DISPLAYS A SYMMETRY RELATED TO THAT OF ICE.  
CC -!- SIMILARITY: Belongs to the bacterial ice nucleation protein  
family.  
CC  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
between the Swiss Institute of Bioinformatics and the EMBL outstation -  
the European Bioinformatics Institute. There are no restrictions on its  
use by non-profit institutions as long as its content is in no way  
modified and this statement is not removed. Usage by and for commercial  
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC  
CC EMBL; M26382; AAA24823.1; --  
DR PIR; JQ0188; JQ0188.  
DR HSSP; P06620; 11NA.  
DR InterPro; IPR000258; Ice nucleatn.  
DR Pfam; PF00818; Ice nucleation; 65.  
DR PRINTS; PR00327; ICNNUCLEATN.  
DR PROSITE; PS00314; ICE NUCLEATION; 45.  
KW Ice nucleation; Repeat; Outer membrane.  
FT DOMAIN 162 1217;  
FT SEQUENCE 1258 AA; 125084 MW; 590E8A130077FBD4 CRC64;  
Query Match 11.3%; Score 87.5; DB 1; Length 1258;  
Best Local Similarity 25.2%; Pred. No. 8.2;  
Matches 41; Conservative 18; Mismatches 41; Indels 63; Gaps 9;  
QY 12 IVVSGSALAGVVPQW--GGGNGHNGGNS-----GPGSTLSIYQYGSNAAL 57  
Db 161 IATYGTSLGTHOSQLIAGYSGTETAGDSSTLIAGYSGTGTAGADSTL-VAGYGTQT- 218  
QY 58 ALQSDARKSETTITQSGYNGADYDQLVTRVVTHEMAHFRNNAITDWNNAKNSDITVQ 117  
Db 219 -----GESSOMAGYGS-----TQGMKGSDLTAG- 243  
QY 118 YG-----GNNALV-----NOTASDSSVMVRQVGFNNATANQ 150  
Db 244 YGSTGTAGDSSLIAGYSGTGTAGDSLSLT--AGYSGTQTAAQ 284  
RESULT 11  
CSGB\_SALTI STANDARD; PRT; 151 AA.  
AC Q8Z7M3;  
DT 28-FEB-2003 (Rel. 41, Created)  
DT 28-FEB-2003 (Rel. 41, Last sequence update)  
DT 10-OCT-2003 (Rel. 42, Last annotation update)  
DE Minor curlin subunit precursor.  
GN CSGB OR STYL180 OR TL177.

OS Salmonella typhi.  
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;  
OC Enterobacteriaceae; Salmonella.  
OX NCBI\_TaxID=601;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=CT18;  
RX MEDLINE=21534947; PubMed=11677608;  
RA Parkhill J., Dougan G., James K.D., Thomson N.R., Pickard D., Wain J.,  
Churcher C., Mungall K.L., Bentley S.D., Holden M.T.G., Sebatia M.,  
Baker S., Basham D., Brooks K., Chillingworth T., Connerton P.,  
Cronin A., Davis P., Davies R.M., Dowd L., White N., Farrar J.,  
Feltwell T., Hamlin N., Haque A., Hien T.T., Holroyd S., Jagels K.,  
Krogh A., Larsen T.S., Leather S., Moule S., O'Gaora P., Parry C.,  
Quail M.A., Rutherford K., Simmonds M., Skelton J., Stevens K.,  
Whitehead S., Barrall B.G.;  
RT "Complete genome sequence of a multiple drug resistant Salmonella  
Nature 413:848-852(2001)."  
RL Nature 413:848-852(2001).  
RN [2]  
RP SEQUENCE FROM N.A.  
RC STRAIN=Ty2 / ATCC 700931;  
RX MEDLINE=22531367; PubMed=12644504;  
RA Deng W., Liou S.-R., Plunkett G. III, Mayhew G.F., Rose D.J.,  
Burland V., Kodyanni V., Schwartz D.C., Blattner F.R.;  
RT "Comparative genomics of Salmonella enterica serovar Typhi strains Ty2  
and CT18.";  
RL J. Bacteriol. 185:2330-2337(2003).  
CC -!- FUNCTION: CURLIN IS THE STRUCTURAL SUBUNIT OF THE CURLI. CURLI ARE  
COILED SURFACE STRUCTURES THAT ASSEMBLE PREFERENTIALLY AT GROWTH  
TEMPERATURES BELOW 37 DEGREES CELSIUS. CURLI CAN BIND TO  
FIBRONECTIN. THE MINOR SUBUNIT IS THE NUCLEATION COMPONENT OF  
CURLIN MONOMERS.  
CC -!- SIMILARITY: BELONGS TO THE CSGA/CSGB FAMILY.  
CC  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
between the Swiss Institute of Bioinformatics and the EMBL outstation -  
the European Bioinformatics Institute. There are no restrictions on its  
use by non-profit institutions as long as its content is in no way  
modified and this statement is not removed. Usage by and for commercial  
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC  
CC EMBL; AL627269; CAD08267.1; --  
DR EMBL; AE016840; AAC069400.1; --  
KW Fimbria; Signal; Complete proteome.  
FT SIGNAL 1 21  
FT CHAIN 22 151 MINOR CURLIN SUBUNIT.  
FT SEQUENCE 151 AA; 16254 MW; 161C54326E573495 CRC64;  
Query Match 11.2%; Score 86.5; DB 1; Length 151;  
Best Local Similarity 26.2%; Pred. No. 0.89;  
Matches 34; Conservative 17; Mismatches 42; Indels 37; Gaps 4;  
QY 10 AAVVSGSALAGVVPQWGGGNGHNGGNSGPGSTLSIYQYGSNAALALQSDARKSETT 69  
Db 58 ARVRQEGSKLLSVISQ--EGENNRKVDQAGNFAVIEQTGNAN-----DAS 103  
QY 70 ITQSGYNGADYDQLVTRVVTHEMAHFRNNAITDWNNAKNSDITVQYGGNNAALVNQT 129  
Db 104 ISQAYGNSA-----AIIQKSGNKANIT--QYGTQKTAVVQK 140  
QY 130 ASDSSVMVRQ 139  
Db 141 QSHMAIRVTQ 150  
RESULT 12  
GP63\_LEIDO STANDARD; PRT; 590 AA.  
ID GP63\_LEIDO  
AC P23223;  
DT 01-NOV-1991 (Rel. 20, Created)  
DT 01-NOV-1991 (Rel. 20, Last sequence update)

10-OCT-2003 (Rel. 42, Last annotation update)  
 DE Leishmanolysin precursor (EC 3.4.24.36) (Cell surface protease)  
 DE (Major surface glycoprotein) (GP63 protein) (Promastigote surface  
 DE endopeptidase).  
 GN GP63  
 OS Leishmania donovani.  
 OC Eukaryota; Euglenozoa; Kinetoplastida; Trypanosomatidae; Leishmania.  
 OX NCBI\_TaxID=5661;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=LV9;  
 RX MEDLINE=92107220; PubMed=1762629;  
 RA Webb J.R., Button L.L., McMaster R.W.;  
 RT "Heterogeneity of the genes encoding the major surface glycoprotein  
 of Leishmania donovani.";  
 RL Mol. Biochem. Parasitol. 48:173-184 (1991).  
 CC -!- FUNCTION: Has an integral role during the infection of macrophages  
 in the mammalian host.  
 CC -!- CATALYTIC ACTIVITY: Preference for hydrophobic residues at P1 and  
 P1' and basic residues at P2 and P3'. A model nonapeptide is  
 CC cleaved at -Ala-Tyr-[Leu-Lys-Lys-  
 CC -!- COFACTOR: Binds 1 zinc ion per subunit (By similarity).  
 CC -!- SUBCELLULAR LOCATION: Attached to the membrane by a GPI-anchor.  
 CC -!- SIMILARITY: Belongs to peptidase family M8.  
 CC  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC  
 CC EMBL; M60048; AAA29244.1; -;  
 CC HSSP; P08148; 1LML.  
 CC MEROPS; M08.001; -;  
 CC InterPro; IPR006025; Pept\_M\_Zn\_BS.  
 CC InterPro; IPR001577; Peptidase\_M8.  
 CC Pfam; PF01457; Peptidase\_M8; 1.  
 CC PRINTS; PR00782; LSHMANOLYSIN.  
 CC PROSITE; PS00142; ZINC\_PROTEASE; 1.  
 CC Hydrolase; Metalloprotease; Glycoprotein; Metal-binding; Zinc;  
 KW Zymogen; Signal; Cell adhesion; GPI-anchor; Lipoprotein.  
 FT SIGNAL 1 39 POTENTIAL.  
 FT PROPEP 40 87 ACTIVATION PEPTIDE.  
 FT CHAIN 68 565 LEISHMANOLYSIN.  
 FT PROPEP 566 590 REMOVED IN NATURE FORM (BY SIMILARITY).  
 FT METAL 251 251 ZINC (CATALYTIC) (BY SIMILARITY).  
 FT ACT\_SITE 252 252  
 FT METAL 255 255 ZINC (CATALYTIC) (BY SIMILARITY).  
 FT METAL 321 321 ZINC (CATALYTIC) (BY SIMILARITY).  
 FT DISULFID 112 129 BY SIMILARITY.  
 FT DISULFID 178 217 BY SIMILARITY.  
 FT DISULFID 301 373 BY SIMILARITY.  
 FT DISULFID 380 443 BY SIMILARITY.  
 FT DISULFID 393 412 BY SIMILARITY.  
 FT DISULFID 402 477 BY SIMILARITY.  
 FT DISULFID 454 498 BY SIMILARITY.  
 FT DISULFID 503 553 BY SIMILARITY.  
 FT DISULFID 523 546 BY SIMILARITY.  
 FT CARBOHYD 287 287 N-LINKED (GLCNAC...) (POTENTIAL).  
 FT LIPID 565 565 GPI-anchor amidated asparagine (By  
 similarity).  
 SQ SEQUENCE 590 AA; 62950 MW; 0FB315D299659F58 CRC64;  
 Query Match 11.1%; Score 86; DB 1; Length 590;  
 Best Local Similarity 48.8%; Pred. No. 4.7;  
 Matches 21; Conservative 5; Mismatches 9; Indels 8; Gaps 2;  
 QY 81 YDQLVTRVVTTHMAHA-----FRNATIDQ-WNAKNSDIT 115  
 |||||  
 Db 241 YDQLVTRVVTTHMAHALGFSVVFRRDARILEISNVHRKDFV 283

RESULT 13  
 GP63 LEIME STANDARD; PRT; 646 AA.  
 AC P43150;  
 DT 01-NOV-1995 (Rel. 32, Created)  
 DT 01-NOV-1995 (Rel. 42, Last sequence update)  
 DT 10-OCT-2003 (Rel. 42, Last annotation update)  
 DE Leishmanolysin C1 precursor (EC 3.4.24.36) (Cell surface protease)  
 DE (Major surface glycoprotein) (GP63 protein) (Promastigote surface  
 DE endopeptidase).  
 GN GP63-C1.  
 OS Leishmania mexicana.  
 OC Eukaryota; Euglenozoa; Kinetoplastida; Trypanosomatidae; Leishmania.  
 OX NCBI\_TaxID=5665;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=MNYC/BZ/62/M379;  
 RX MEDLINE=93149206; PubMed=8426614;  
 RA Medina-Acosta E., Kares R.E., Russell D.G.;  
 RT "Structurally distinct genes for the surface protease of Leishmania  
 mexicana are developmentally regulated.";  
 RL Mol. Biochem. Parasitol. 57:31-46 (1993).  
 CC -!- FUNCTION: Has an integral role during the infection of macrophages  
 in the mammalian host.  
 CC -!- CATALYTIC ACTIVITY: Preference for hydrophobic residues at P1 and  
 P1' and basic residues at P2 and P3'. A model nonapeptide is  
 CC cleaved at -Ala-Tyr-[Leu-Lys-Lys-  
 CC -!- COFACTOR: Binds 1 zinc ion per subunit (By similarity).  
 CC -!- DEVELOPMENTAL STAGE: Expressed in both the promastigote and the  
 CC amastigote forms.  
 CC -!- SIMILARITY: Belongs to peptidase family M8.  
 CC  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC  
 CC EMBL; X64394; CAA45733.1; -;  
 CC F1R; S19916; S19916.  
 CC HSSP; P08148; 1LML.  
 CC MEROPS; M08.001; -;  
 CC GlycoSuiteDB; P43150; -;  
 CC InterPro; IPR006025; Pept\_M\_Zn\_BS.  
 CC InterPro; IPR001577; Peptidase\_M8.  
 CC Pfam; PF01457; Peptidase\_M8; 1.  
 CC PRINTS; PR00782; LSHMANOLYSIN.  
 CC PROSITE; PS00142; ZINC\_PROTEASE; 1.  
 CC Hydrolase; Metalloprotease; Glycoprotein; Metal-binding; Zinc;  
 KW Zymogen; Signal; Cell adhesion; Multigene family.  
 FT SIGNAL 1 39 POTENTIAL.  
 FT PROPEP 40 102 ACTIVATION PEPTIDE (POTENTIAL).  
 FT CHAIN 103 646 LEISHMANOLYSIN C1.  
 FT METAL 266 266 ZINC (CATALYTIC) (BY SIMILARITY).  
 FT ACT\_SITE 267 267 BY SIMILARITY.  
 FT METAL 270 270 ZINC (CATALYTIC) (BY SIMILARITY).  
 FT METAL 336 336 ZINC (CATALYTIC) (BY SIMILARITY).  
 FT DISULFID 127 144 BY SIMILARITY.  
 FT DISULFID 193 232 BY SIMILARITY.  
 FT DISULFID 316 388 BY SIMILARITY.  
 FT DISULFID 395 458 BY SIMILARITY.  
 FT DISULFID 408 427 BY SIMILARITY.  
 FT DISULFID 417 492 BY SIMILARITY.  
 FT DISULFID 469 513 BY SIMILARITY.  
 FT DISULFID 518 568 BY SIMILARITY.  
 FT DISULFID 538 561 BY SIMILARITY.  
 FT CARBOHYD 86 86 N-LINKED (GLCNAC...) (POTENTIAL).  
 FT CARBOHYD 297 297 N-LINKED (GLCNAC...) (POTENTIAL).  
 FT CARBOHYD 399 399 N-LINKED (GLCNAC...) (POTENTIAL).  
 FT CARBOHYD 409 409 N-LINKED (GLCNAC...) (POTENTIAL).

FT CARBOHYD 433 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 445 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 466 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 501 N-LINKED (GLCNAC. . .) (POTENTIAL).  
SQ SEQUENCE 646 AA; 59054 MW; FE448DDC78C10B0A CRC64;  
  
Query Match 10.9%; Score 84.5; DB 1; Length 646;  
Best Local Similarity 36.5%; Pred. No. 6.9;  
Matches 23; Conservative 7; Mismatches 12; Indels 21; Gaps 2;  
  
QY 81 YDQVTRVTVTHMAFAFRNNR-----IDWNNAKNSDITV-----GOYG 119  
DB 256 YDQVTRVTVTHMAFAFRNNR-----IDWNNAKNSDITV-----GOYG 119  
  
QY 120 GNN 122  
DB 316 CNS 318  
  
RESULT 14  
GP63 LEICH STANDARD; PRT; 599 AA.  
AC P15706;  
DT 01-APR-1990 (Rel. 14, Created)  
DT 01-APR-1990 (Rel. 14, Last sequence update)  
DT 10-OCT-2003 (Rel. 42, Last annotation update)  
DE Leishmanolysin precursor (EC 3.4.24.36) (Cell surface protease)  
DE (Major surface glycoprotein) (GP63 protein) (Promastigote surface  
DE endopeptidase).  
GN GP63.  
OS Leishmania chagasi.  
OC Eukaryota; Euklenozoa; Kinetoplastida; Trypanosomatidae; Leishmania.  
OX NCBI\_TaxID=44271;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=92112918; PubMed=1370484;  
RA Ramamoorthy R., Donelson J.E., Paetz K.E., Maybodi M., Roberts S.C.,  
RA Wilson M.E.  
RT "Three distinct RNAs for the surface protease gp63 are differentially  
RT expressed during development of Leishmania donovani chagasi  
RT promastigotes to an infectious form."  
RL J. Biol. Chem. 267:1888-1895(1992).  
CC -!- FUNCTION: Has an integral role during the infection of macrophages  
CC in the mammalian host.  
CC -!- CATALYTIC ACTIVITY: Preference for hydrophobic residues at P1 and  
CC P1', and basic residues at P2 and P3'. A model nonapeptide is  
CC cleaved at -Ala-Tyr-|-Leu-Lys-Lys-.  
CC -!- COFACTOR: Binds 1 zinc ion per subunit (By similarity).  
CC -!- SUBCELLULAR LOCATION: Attached to the membrane by a GPI-anchor.  
CC -!- SIMILARITY: Belongs to peptidase family M8.  
CC  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC  
CC ENBL; M80672; AAA29238.1;  
CC ENBL; M28527; AAA29235.1;  
CC PIR; A44951; A44951.  
CC HSP; P08148; 1LML.  
CC MEROPS; M08.001;  
CC InterPro; IPR006025; Pept M.Zn.BS.  
CC InterPro; IPR001577; Peptidase\_M8.

DR PFam; PF01457; Peptidase\_M8; 1.  
DR PRINTS; PS00782; LSHMANOLYSIN.  
DR PROSITE; PS00142; ZINC\_PROTEASE; 1.  
KW Hydrolase; Metalloprotease; Glycoprotein; Metal-binding; Zinc;  
KW Zymogen; Signal; Cell adhesion; GPI-anchor; Lipoprotein.  
FT SIGNAL 1 39 POTENTIAL.  
FT PROPEP 40 97 ACTIVATION PEPTIDE.  
FT CHAIN 98 574. LEISHMANOLYSIN.  
FT PROPEP 575 599 REMOVED IN MATURE FORM (BY SIMILARITY).  
FT METAL 261 261 ZINC (CATALYTIC) (BY SIMILARITY).  
FT ACT SITE 262 262 BY SIMILARITY.  
FT METAL 265 265 ZINC (CATALYTIC) (BY SIMILARITY).  
FT METAL 331 331 ZINC (CATALYTIC) (BY SIMILARITY).  
FT DISULFID 122 139 BY SIMILARITY.  
FT DISULFID 188 227 BY SIMILARITY.  
FT DISULFID 311 383 BY SIMILARITY.  
FT DISULFID 390 452 BY SIMILARITY.  
FT DISULFID 403 422 BY SIMILARITY.  
FT DISULFID 412 486 BY SIMILARITY.  
FT DISULFID 463 507 BY SIMILARITY.  
FT DISULFID 512 562 BY SIMILARITY.  
FT DISULFID 532 555 BY SIMILARITY.  
FT CARBOHYD 297 297 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 394 394 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT LIPID 574 574 GPI-anchor amidated asparagine (By  
FT similarity).  
SQ SEQUENCE 599 AA; 63848 MW; 746730AE8E2A2E7C CRC64;  
  
Query Match 10.8%; Score 84; DB 1; Length 599;  
Best Local Similarity 37.5%; Pred. No. 7;  
Matches 24; Conservative 5; Mismatches 29; Indels 6; Gaps 1;  
  
QY 39 SGPDSTLSIQYGSANAALA-----LQSDARKSETTITQSGYNGADYDQVTRVTVTHE 92  
DB 203 SNTDFVMYVASVSEEGVLAWATTCQVFSGDHPGVGINIPAAINASRYDQVTRVTVTHE 262  
QY 93 MAHA 96  
DB 263 MAHA 266  
  
RESULT 15  
HLIA SERMA STANDARD; PRT; 1608 AA.  
AC P15320;  
DT 01-APR-1990 (Rel. 14, Created)  
DT 01-APR-1990 (Rel. 14, Last sequence update)  
DT 01-NOV-1990 (Rel. 16, Last annotation update)  
DE Hemolysin precursor.  
GN SHLA.  
OS Serratia marcescens.  
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;  
OC Enterobacteriaceae; Serratia.  
OX NCBI\_TaxID=615;  
RN [1]  
RP SEQUENCE FROM N.A., AND SEQUENCE OF 31-40.  
RC STRAIN=SR8; PubMed=3290200;  
RX MEDLINE=88257037; PubMed=3290200;  
RA Poole K., Schiebel E., Braun V.;  
RT "Molecular characterization of the hemolysin determinant of Serratia  
RT marcescens."  
RL J. Bacteriol. 170:3177-3188(1988).  
CC -!- FUNCTION: Bacterial hemolysins are exotoxins that attack blood  
CC cell membranes and cause cell rupture by mechanisms not clearly  
CC defined.  
CC -!- FUNCTION: CELL-BOUND HEMOLYSIN, WHICH RELEASES HEME-IRON FROM  
CC ERYTHROCYTES BY INTERACTION WITH THE ERYTHROCYTE MEMBRANE. SHLA  
CC REQUIRES SHLB FUNCTION.  
CC -!- SUBCELLULAR LOCATION: Outer membrane.  
CC -!- SIMILARITY: TO P.MIRABILIS HEMOLYSIN (HPMA).  
CC  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -



This Page Blank (uspto)

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: August 2, 2004, 14:39:17 ; Search time 29.7 Seconds  
(without alignments)  
1604.150 Million cell updates/sec

Title: US-09-543-407-28  
Perfect score: 775  
Sequence: 1 MKLLKVAFAAIVVSGSALA.....DSSVMVQVFGNNATANQY 151

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 1017041 seqs, 315518202 residues  
Total number of hits satisfying chosen parameters: 1017041

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000  
Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : SPTREMBL\_25:  
1: sp\_archaea:  
2: sp\_bacteria:  
3: sp\_fungi:  
4: sp\_human:  
5: sp\_invertebrate:  
6: sp\_mammal:  
7: sp\_mhc:  
8: sp\_organelle:  
9: sp\_phage:  
10: sp\_plant:  
11: sp\_rodent:  
12: sp\_virus:  
13: sp\_vertebrate:  
14: sp\_unclassified:  
15: sp\_rviro:  
16: sp\_bacterioph:  
17: sp\_archaeap:

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	ID	Description
1	674	87.0	152	033802	O33802 salmonella
2	566.5	73.1	150	07X243	Q7X243 citrobacter
3	530	68.4	149	07X240	Q7X240 citrobacter
4	484.5	62.5	152	08CW63	Q8CW63 escherichia
5	411.5	53.1	150	07X237	Q7X237 enterobacte
6	295	38.1	76	Q54069	Q54069 salmonella
7	122	15.7	29	Q9S3J5	Q9S3J5 escherichia
8	112	14.5	130	08SJI4	Q8SJI4 bradyrhizob
9	111.5	14.4	3501	08XSD6	Q8XSD6 ralstonia s
10	111.5	14.4	3552	08XSD6	Q8XSD6 ralstonia s
11	110	14.2	139	08EIH3	Q8EIH3 shewanella
12	108	13.9	1422	08EFU3	Q8EFU3 shewanella
13	107	13.8	2174	092U08	Q92U08 rhizobium m
14	106.5	13.7	502	08EIH4	Q8EIH4 shewanella
15	99.5	12.8	171	08SJI3	Q8SJI3 bradyrhizob
16	99	12.8	714	07U5X6	Q7U5X6 synechococc

17	98	12.6	179	2	033801	O33801 salmonella
18	97.5	12.6	154	16	Q89JIS	Q89JIS bradyrhizob
19	97	12.5	1748	5	Q94821	Q94821 tetrahymena
20	95	12.3	368	16	Q8SMD6	Q8SMD6 mycoplasma
21	95	12.3	400	10	Q8LAR8	Q8LAR8 oryza sativ
22	95	12.3	1410	16	Q8CMJ0	Q8CMJ0 shewanella
23	94.5	12.2	160	16	Q8CW64	Q8CW64 escherichia
24	94	12.1	362	16	Q8EV84	Q8EV84 mycoplasma
25	94	12.1	645	16	Q7UIC5	Q7UIC5 mycobacteri
26	94	12.1	646	16	O53818	O53818 mycobacteri
27	93.5	12.1	644	5	Q8MNY9	Q8MNY9 leishmania
28	93	12.0	157	16	Q88HG0	Q88HG0 pseudomonas
29	92.5	11.9	348	13	Q93397	Q93397 cyprinus ca
30	92.5	11.9	644	5	O43994	O43994 leishmania
31	92.5	11.9	1765	16	Q7V8S5	Q7V8S5 prochloroco
32	92	11.9	598	5	Q25275	Q25275 leishmania
33	91	11.7	152	2	Q7X241	Q7X241 citrobacter
34	91	11.7	573	10	Q9SAF2	Q9SAF2 arabidopsis
35	91	11.7	614	16	Q7TYR8	Q7TYR8 mycobacteri
36	91	11.7	615	16	P95249	P95249 mycobacteri
37	91	11.7	1408	16	Q8E833	Q8E833 shewanella
38	90.5	11.7	1286	2	Q841Y5	Q841Y5 campylobact
39	89.5	11.5	151	2	Q7X238	Q7X238 enterobacte
40	89.5	11.5	151	16	Q7UCZ1	Q7UCZ1 shigella fl
41	89.5	11.5	153	16	Q89JI6	Q89JI6 bradyrhizob
42	89.5	11.5	160	16	Q83RU7	Q83RU7 shigella fl
43	89.5	11.5	1333	16	Q8FD38	Q8FD38 xanthomonas
44	89.5	11.5	2411	16	Q8PF72	Q8PF72 xanthomonas
45	89	11.5	354	16	P95248	P95248 mycobacteri

ALIGNMENTS

RESULT 1

O33802	PRELIMINARY;	PRT;	152 AA.
ID	O33802		
AC	O33802;		
DT	01-JAN-1998 (TREMBLrel. 05, Created)		
DT	01-JAN-1998 (TREMBLrel. 05, Last sequence update)		
DT	01-DEC-2001 (TREMBLrel. 19, Last annotation update)		
DE	AgfA protein (Fragment).		
GN	AGFA.		
OS	Salmonella typhimurium.		
OC	Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;		
OC	Enterobacteriaceae; Salmonella.		
OX	NCBI_TaxID=602;		
RN	[1]		
RP	SEQUENCE FROM N.A.		
RX	MEDLINE=98053981; PubMed=9393832;		
RA	Sukupolvi S.S., Lorentz R.G., Gordon J.I., Bian Z., Pfeifer J.D.,		
RA	Normark S.J., Rhen M.;		
RT	"Expression of thin, aggregative fimbriae promotes interaction of		
RT	Salmonella typhimurium SR-11 with mouse small intestinal epithelial		
RT	cells."		
RL	Infect. Immun. 65:5320-5325(1997).		
DR	EMBL; AJ000514; CAA04151.1; -		
FT	NON TER 152		
SQ	SEQUENCE 152 AA; 15401 MW; 9DA7DADC2364B006 CRC64;		

Query Match 87.0%; Score 674; DB 2; Length 152;  
Best Local Similarity 88.7%; Pred. No. 1.1e-47;  
Matches 134; Conservative 2; Mismatches 15; Indels 0; Gaps 0;

Qy	1	MKLLKVAFAAIVVSGSALA	QVQWGGGNGHNGSSGPDSTLSIYQGSANAALQ	60
Db	1	MKLLKVAFAAIVVSGSALA	QVQWGGGNGHNGSSGPDSTLSIYQGSANAALQ	60
Qy	61	SDARKSETTITQSGYNGADYDQLVTRVVTTHMAHAFRNATIDQWNAKNSDITVGYGG	120	
Db	61	SDARKSETTITQSGYNGADYDQADNSITELTQNGFRNATIDQWNAKNSDITVGYGG	120	
Qy	121	NNAALVNQTASDSSVMVQVFGNNATANQY	151	

```

Db      121 NNAALVNOTASDSSVMVROVQVGFNNAPANQY 151
|||||
RESULT 2
Q7X243
ID Q7X243 PRELIMINARY; PRT; 150 AA.
AC Q7X243;
DT 01-OCT-2003 (T-EMBLrel. 25, Created)
DT 01-OCT-2003 (T-EMBLrel. 25, Last sequence update)
DT 01-OCT-2003 (T-EMBLrel. 25, Last annotation update)
DE Curlin-csgA protein.
GN CSGA.
OS Citrobacter sp. Pec2.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Citrobacter.
OX NCBI_TaxID=213763;
RN [1]
RC STRAIN=Pec2.
RA Zogaj X., Bokranz W., Nintz M., Romling U.;
RT "Production of Cellulose and Curli Fimbriae by Members of the Family
RL Enterobacteriaceae Isolated from the Human Gastrointestinal Tract.";
RL Infect. Immun. 72:4151-4158(2003).
DR EMBL; AJ515700; CAD56672.1; -.
SQ SEQUENCE 150 AA; 15016 MW; 1D7141B8D6973DC6 CRC64;

Query Match 73.1%; Score 566.5; DB 2; Length 150;
Best Local Similarity 76.8%; Pred. No. 6.5e-39;
Matches 116; Conservative 10; Mismatches 24; Indels 1; Gaps 1;

QY 1 MKLLKVAAPAAIVVSGSALAGVVPQW-GGGGNGGNGSGGPDSTLSIYQYGSANAALALQ 60
Db 1 MKLLQVAAAPAAIVVSGSALAGVVPQWGGGG-GGGSSSGGPSTLSIYQYGSANAALALQ 59
QY 61 SDARKSETTITQSGYNGADYDQLVTRVVTHEMAHAFRNATIDQWNAKNSDITVGYGG 120
Db 60 SDARKSDTTIHQGFNGADVGQSGDSTIDLTQNGFKNNATIDQWNGKNSDITVGYGG 119
QY 121 NNAALVNOTASDSSVMVROVQVGFNNATANQY 151
Db 120 HNAALVNOTASDSSVMVROVQVGFNNATANQY 150

RESULT 3
Q7X240
ID Q7X240 PRELIMINARY; PRT; 149 AA.
AC Q7X240;
DT 01-OCT-2003 (T-EMBLrel. 25, Created)
DT 01-OCT-2003 (T-EMBLrel. 25, Last sequence update)
DT 01-OCT-2003 (T-EMBLrel. 25, Last annotation update)
DE Curlin-csgA protein.
GN CSGA.
OS Citrobacter freundii.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Citrobacter.
OX NCBI_TaxID=546;
RN [1]
RC STRAIN=Pec4;
RA Zogaj X., Bokranz W., Nintz M., Romling U.;
RT "Production of Cellulose and Curli Fimbriae by Members of the Family
RL Enterobacteriaceae Isolated from the Human Gastrointestinal Tract.";
RL Infect. Immun. 72:4151-4158(2003).
DR EMBL; AJ515701; CAD56675.1; -.
SQ SEQUENCE 149 AA; 15260 MW; 946DD52017F648FD CRC64;

Query Match 68.4%; Score 530; DB 2; Length 149;
Best Local Similarity 71.5%; Pred. No. 6.3e-36;
Matches 108; Conservative 16; Mismatches 25; Indels 2; Gaps 1;

QY 1 MKLLKVAAPAAIVVSGSALAGVVPQWGGGNGGNGSGGPDSTLSIYQYGSANAALALQ 60
|||||

```

```

Db      1 MKLLKVAAPAAIVVSGSALAGVVPQW--GGHHGGGSGNYGPDSSLSIYQYGSNSANAALQ 58
QY 61 SDARKSETTITQSGYNGADYDQLVTRVVTHEMAHAFRNATIDQWNAKNSDITVGYGG 120
Db 59 SDARKSDTTITQHGRNGAVVVGQAGDDSTISLQGTQFQNSATIDQWNAKNSDITVGYGG 118
QY 121 NNAALVNOTASDSSVMVROVQVGFNNATANQY 151
Db 119 RUGALVNOTASDSSVMVROVQVGFNNATANQY 149

RESULT 4
Q8CW63
ID Q8CW63 PRELIMINARY; PRT; 152 AA.
AC Q8CW63;
DT 01-MAR-2003 (T-EMBLrel. 23, Created)
DT 01-MAR-2003 (T-EMBLrel. 23, Last sequence update)
DT 01-MAR-2003 (T-EMBLrel. 23, Last annotation update)
DE Major curlin subunit precursor.
GN CSGA OR C1306.
OS Escherichia coli O6.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Escherichia.
OX NCBI_TaxID=217992;
RN [1]
RC SEQUENCE FROM N.A.
RA STRAIN=O6:H1 / CFT073 / ATCC 700928;
RX MEDLINE=2238234; PubMed=12471157;
RA Welch R.A., Burland V., Plunkett G. III, Redford P., Roesch P.,
RA Rasko D., Buckles E.L., Liou S.-R., Boutin A., Hackett J., Stroud D.,
RA Mayhew G.F., Rose D.J., Zhou S., Schwartz D.C., Perna N.T.,
RA Mobley H.L.T., Donnenberg M.S., Blattner F.R.;
RT "Extensive mosaic structure revealed by the complete genome sequence
of uropathogenic Escherichia coli.";
RL Proc. Natl. Acad. Sci. U.S.A. 99:17020-17024(2002).
DR EMBL; AE016759; AAN79779.1; -.
RW Complete proteome.
SQ SEQUENCE 152 AA; 15064 MW; 3BA57F34C1240B83 CRC64;

Query Match 62.5%; Score 484.5; DB 16; Length 152;
Best Local Similarity 66.4%; Pred. No. 3.4e-32;
Matches 101; Conservative 15; Mismatches 35; Indels 1; Gaps 1;

QY 1 MKLLKVAAPAAIVVSGSALAGVVPQW-GGGGNGGNGSGGPDSTLSIYQYGSANAALAL 59
Db 1 MKLLKVAAPAAIVVSGSALAGVVPQWGGGGNGGNGSGGPDSTLSIYQYGSANAALAQ 60
QY 60 QSDARKSETTITQSGYNGADYDQLVTRVVTHEMAHAFRNATIDQWNAKNSDITVGYGG 119
Db 61 QADARNSDLTITQHGNGADVGQSGDSSIDLTQGFNGSATLDQWNGKSDTMTVKQFG 120
QY 120 GNNALVNOTASDSSVMVROVQVGFNNATANQY 151
Db 121 GNGAAVDOTASNSVNVTVQVGFNNATAFY 152

RESULT 5
Q7X237
ID Q7X237 PRELIMINARY; PRT; 150 AA.
AC Q7X237;
DT 01-OCT-2003 (T-EMBLrel. 25, Created)
DT 01-OCT-2003 (T-EMBLrel. 25, Last sequence update)
DT 01-OCT-2003 (T-EMBLrel. 25, Last annotation update)
DE Curlin-csgA protein.
GN CSGA.
OS Enterobacter sakazakii.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Enterobacter.
OX NCBI_TaxID=28141;
RN [1]
RC SEQUENCE FROM N.A.
RA STRAIN=Pec39;
RA Zogaj X., Bokranz W., Nintz M., Romling U.;

```



RT "Production of Cellulose and Curli Fimbriae by Members of the Family  
RL Enterobacteriaceae Isolated from the Human Gastrointestinal Tract."  
RL Infect. Immun. 72:4151-4158(2003).  
DR EMBL: AJ515702; CAD56678.1; -  
SQ SEQUENCE 150 AA; 15112 MW; 5D8BB2D872DF15F3 CRC64;

Query Match 53.1%; Score 411.5; DB 2; Length 150;  
Best Local Similarity 57.8%; Pred. No. 3.1e-26;  
Matches 87; Conservative 25; Mismatches 38; Indels 1; Gaps 1;

QY 1 MKLLKVAFAAIVVSGSALAGVVPQWGGNGGNSGSPDSTLSIYQYGSANAALALQ 60  
DB 1 MKFIKVAALAAIVVSSAMAGNLQ-GGWGHGHGGYGPNSLNIYQNGGNSALALQ 59

QY 61 SDARKSETTITQSGYNGADYDQLVTRVTHEMAHAFRNATIDQNAKNSDITVQYGG 120  
DB 60 TDARNVNLISQTGGGNGADVGQGSDDSSINTQNGFGNSATIDQNSKDSVMNVSYQGG 119

QY 121 NNAALVNOTASDSSVMVROVGFNNATANQY 151  
DB 120 LNCALVDQTASNSTVNTQIGFGNHATAHQY 150

RESULT 6  
Q54069 PRELIMINARY; PRT; 76 AA.  
AC Q54069  
DT 01-NOV-1996 (T-EMBLrel. 01, Created)  
DT 01-NOV-1996 (T-EMBLrel. 01, Last sequence update)  
DT 01-DEC-2001 (T-EMBLrel. 19, Last annotation update)  
DE SEF17 fimbria (Fragment).  
GN AGFA.  
OS Salmonella enteritidis.  
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;  
OC Enterobacteriaceae; Salmonella.  
OX NCBI\_TaxID=592;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=SE30;  
RA Cox J.M., Eglezos S., Woolcock J.B.;  
RT "Virulence of salmonella enteritidis in chickens correlates with  
RT colony morphology and expression of SEF17 fimbriae."  
RL Submitted (APR-1996) to the EMBL/GenBank/DBJ databases.  
DR EMBL: U53207; AAA98671.1; -  
FT NON TER 1  
FT NON TER 1  
FT NON TER 76  
SQ SEQUENCE 76 AA; 7704 MW; 2FD5411241A7BCB1 CRC64;

Query Match 38.1%; Score 295; DB 2; Length 76;  
Best Local Similarity 78.9%; Pred. No. 4.7e-17;  
Matches 60; Conservative 1; Mismatches 15; Indels 0; Gaps 0;

QY 30 GNHNGGNSGSPDSTLSIYQYGSANAALALQSDARKSETTITQSGYNGADYDQLVTRV 89  
DB 1 GNHXGGNSGSPDSTLSIYQYGSANAALALQSDARKSETTITQSGYNGADYDQGDNST 60

QY 90 THEMAHAFRNATIDQ 105  
DB 61 IELTQNGFRNATIDQ 76

RESULT 7  
Q9S3J5 PRELIMINARY; PRT; 29 AA.  
AC Q9S3J5  
DT 01-MAY-2000 (T-EMBLrel. 13, Created)  
DT 01-MAY-2000 (T-EMBLrel. 13, Last sequence update)  
DT 01-MAR-2001 (T-EMBLrel. 16, Last annotation update)  
DE Curlin subunit monomer (Fragment).  
GN CSGA.  
OS Escherichia coli.  
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;  
OC Enterobacteriaceae; Escherichia.

OX NCBI\_TaxID=562;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC TRANSPOSON=Insertion sequence ISI;  
RX MEDLINE=99314153; PubMed=10386375;  
RA La Ragione R.M., Collighan R.J., Woodward M.J.;  
RT "Non-curliation of Escherichia coli O78:K80 isolates associated with  
RT ISI inserti on in cs8g and reduced persistence in poultry infection."  
RL FEMS Microbiol. Lett. 175:247-253(1999).  
DR EMBL: AJ131756; CAB45380.1; -  
FT NON TER 29  
FT NON TER 29  
SQ SEQUENCE 29 AA; 2789 MW; E290DFC07ABBB243 CRC64;

Query Match 15.7%; Score 122; DB 2; Length 29;  
Best Local Similarity 89.7%; Pred. No. 0.0021;  
Matches 26; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 MKLLKVAFAAIVVSGSALAGVVPQWGGG 29  
DB 1 MKLLKVAALAAIVFSGSALAGVVPQYGGG 29

RESULT 8  
Q89JI4 PRELIMINARY; PRT; 130 AA.  
AC Q89JI4  
DT 01-JUN-2003 (T-EMBLrel. 24, Created)  
DT 01-JUN-2003 (T-EMBLrel. 24, Last sequence update)  
DT 01-JUN-2003 (T-EMBLrel. 24, Last annotation update)  
DE BL15299 protein.  
GN BL15299.  
OS Bradyrhizobium japonicum.  
OC Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;  
OC Bradyrhizobiaceae; Bradyrhizobium.  
OX NCBI\_TaxID=375;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=USDA 110;  
RX MEDLINE=22484998; PubMed=12597275;  
RA Kaneko T., Nakamura Y., Sato S., Minamisawa K., Uchiumi T.,  
RA Sasamoto S., Watanabe A., Iidesawa K., Iriguchi M., Kawashima K.,  
RA Kohara M., Matsumoto M., Shimpo S., Tsuruoka H., Wada T., Yamada M.,  
RA Tabata S.;  
RT "Complete genomic sequence of nitrogen-fixing symbiotic bacterium  
RT Bradyrhizobium japonicum USDA110."  
RL DNA Res. 9:189-197(2002).  
DR EMBL: AP005954; BAC50564.1; -  
KW Complete proteome.  
SQ SEQUENCE 130 AA; 12699 MW; ACFB2D66A48D260F CRC64;

Query Match 14.5%; Score 112; DB 16; Length 130;  
Best Local Similarity 26.5%; Pred. No. 0.085;  
Matches 40; Conservative 28; Mismatches 55; Indels 28; Gaps 6;

QY 4 LKVAFAAIVVSGSALAGVVPQWGGNGGNSGSPDSTLSIYQYGSANAALALQSDA 63  
DB 1 MRITYLVATAIALSALTITVDAQ-----AGNSA-----SVLQFGTTNSSFSIQGS 45

QY 64 RKSETTITQSGYNGADYDQLVTRVTHEMAHAFRNATIDQ-----NAKNSDITVQY 119  
DB 46 TSNNAITLQFCATNTA-----TTLQIGSLTV--NTAVTQGGTTATASNTALT-GQVG 96

QY 120 GNAALVNOTASDSSVMVROVGFNNATANQ 150  
DB 97 GNSSLIGQIGANTAGVQGLGILNGSTILQ 127

RESULT 9  
Q8YI06 PRELIMINARY; PRT; 3501 AA.  
AC Q8YI06  
DT 01-MAR-2002 (T-EMBLrel. 20, Created)  
DT 01-MAR-2002 (T-EMBLrel. 20, Last sequence update)

01-OCT-2003 (TrEMBLrel. 25, Last annotation update)  
DE Probable hemagglutinin-related protein.  
GN RSC0887 OR RS06116.  
OS Ralstonia solanacearum (Pseudomonas solanacearum).  
OC Bacteria; Proteobacteria; Betaproteobacteria; Burkholderiales;  
OC Burkholderiaceae; Ralstonia.  
OX NCBI\_TaxID=305;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=GMI1000;  
RX MEDLINE=21681879; PubMed=11823852;  
RA Salanoubat M., Genin S., Artiguenave F., Gouzy J., Mangenot S.,  
RA Arlat M., Billault A., Brottier P., Camus J.C., Cattolico L.,  
RA Chandler M., Choise N., Claudel-Renard C., Cunnac S., Demange N.,  
RA Caspin C., Lavie M., Moisan A., Robert C., Saurin W., Schiex T.,  
RA Sigquier P., Thebaud P., Whalen M., Wincker P., Levy M.,  
RA Weissenbach J., Boucher C.A.;  
RT "Genome sequence of the plant pathogen Ralstonia solanacearum."  
RL Nature 415:497-502(2002).  
DR EMBL; AL646061; CAD14589.1; --  
DR GO; GO:0004519; F:endonuclease activity; IEA.  
DR GO; GO:0003676; F:nucleic acid binding; IEA.  
DR InterPro; IPR008619; F:haemagg.  
DR InterPro; IPR008638; Haemagg act.  
DR Pfam; PF05594; F:haemagg; 20.  
DR Pfam; PF05860; Haemagg act; 1.  
DR PROSITE; PS01070; NUCLEASE\_NON\_SPEC; 1.  
KW Complete Proteome.  
KW SEQUENCE 3501 AA; 348421 MW; 290B41C99018A107 CRC64;  
Query Match 14.4%; Score 111.5; DB 16; Length 3501;  
Best Local Similarity 29.3%; Pred. No. 5.2;  
Matches 41; Conservative 20; Mismatches 52; Indels 27; Gaps 5;  
QY 15 SGSLAGVVPQGGGNGHNGG-NSSGPDSTLSIYQVGSANAA----- 56  
DB 2431 SGSHFSTAGSWGDLGRNVGGPNSG-----VGLAPYGAHSDNAGNSSRQNASVIG 2486  
QY 57 LALQSDARKSETTITQSGYNGADYDQLVTRVVTHEMAHAFRNATIDQNAKNSDITVG 116  
DB 2487 KSVQVQARTGDIIVSGSGISALSDVLLAKQKGVIVA-----GNDTSSR-HEDHSDRTIG 2541  
QY 117 QYGNNAALVNQTASDSSVM 136  
DB 2542 DLGGNGYSGTVGVRASSTL 2561  
RESULT 10  
ID QXKSD6 PRELIMINARY; PRT; 3552 AA.  
AC QXKSD6;  
DT 01-MAR-2002 (TrEMBLrel. 20, Created)  
DT 01-MAR-2002 (TrEMBLrel. 20, Last sequence update)  
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)  
DE Probable hemagglutinin-related protein.  
GN RSC0887 OR RS06116.  
OS Ralstonia solanacearum (Pseudomonas solanacearum).  
OC Bacteria; Proteobacteria; Betaproteobacteria; Burkholderiales;  
OC Burkholderiaceae; Ralstonia.  
OX NCBI\_TaxID=305;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=GMI1000;  
RX MEDLINE=21681879; PubMed=11823852;  
RA Salanoubat M., Genin S., Artiguenave F., Gouzy J., Mangenot S.,  
RA Arlat M., Billault A., Brottier P., Camus J.C., Cattolico L.,  
RA Chandler M., Choise N., Claudel-Renard C., Cunnac S., Demange N.,  
RA Caspin C., Lavie M., Moisan A., Robert C., Saurin W., Schiex T.,  
RA Sigquier P., Thebaud P., Whalen M., Wincker P., Levy M.,  
RA Weissenbach J., Boucher C.A.;  
RT "Genome sequence of the plant pathogen Ralstonia solanacearum."  
RL Nature 415:497-502(2002).  
DR EMBL; AL646061; CAD14589.1; --  
DR GO; GO:0004519; F:endonuclease activity; IEA.  
DR GO; GO:0003676; F:nucleic acid binding; IEA.  
DR InterPro; IPR008619; F:haemagg.  
DR InterPro; IPR008638; Haemagg act.  
DR Pfam; PF05594; F:haemagg; 20.  
DR Pfam; PF05860; Haemagg act; 1.  
DR PROSITE; PS01070; NUCLEASE\_NON\_SPEC; 1.  
KW Complete Proteome.  
KW SEQUENCE 3501 AA; 348421 MW; 290B41C99018A107 CRC64;  
Query Match 14.4%; Score 111.5; DB 16; Length 3501;  
Best Local Similarity 29.3%; Pred. No. 5.2;  
Matches 41; Conservative 20; Mismatches 52; Indels 27; Gaps 5;  
QY 15 SGSLAGVVPQGGGNGHNGG-NSSGPDSTLSIYQVGSANAA----- 56  
DB 2431 SGSHFSTAGSWGDLGRNVGGPNSG-----VGLAPYGAHSDNAGNSSRQNASVIG 2486  
QY 57 LALQSDARKSETTITQSGYNGADYDQLVTRVVTHEMAHAFRNATIDQNAKNSDITVG 116  
DB 2487 KSVQVQARTGDIIVSGSGISALSDVLLAKQKGVIVA-----GNDTSSR-HEDHSDRTIG 2541  
QY 117 QYGNNAALVNQTASDSSVM 136  
DB 2542 DLGGNGYSGTVGVRASSTL 2561

Nature 415:497-502(2002).  
DR EMBL; AL646079; CAD17691.1; --  
DR GO; GO:0046821; C:extrachromosomal DNA; IEA.  
DR GO; GO:0004519; F:endonuclease activity; IEA.  
DR GO; GO:0003676; F:nucleic acid binding; IEA.  
DR InterPro; IPR001604; Endonuclease.  
DR InterPro; IPR008619; F:haemagg.  
DR InterPro; IPR008638; Haemagg act.  
DR Pfam; PF05594; F:haemagg; 20.  
DR Pfam; PF05860; Haemagg act; 1.  
DR PROSITE; PS01070; NUCLEASE\_NON\_SPEC; 1.  
KW Plasmid; Complete proteome.  
KW SEQUENCE 3552 AA; 352934 MW; C5432AABE2CCF59C CRC64;  
Query Match 14.4%; Score 111.5; DB 16; Length 3552;  
Best Local Similarity 29.3%; Pred. No. 5.3;  
Matches 41; Conservative 20; Mismatches 52; Indels 27; Gaps 5;  
QY 15 SGSLAGVVPQGGGNGHNGG-NSSGPDSTLSIYQVGSANAA----- 56  
DB 2430 SGSHFSTAGSWGDLGRNVGGPNSG-----VGLAPYGAHSDNAGNSSRQNASVIG 2485  
QY 57 LALQSDARKSETTITQSGYNGADYDQLVTRVVTHEMAHAFRNATIDQNAKNSDITVG 116  
DB 2486 KSVQVQARTGDIIVSGSGISALSDVLLAKQKGVIVA-----GNDTSSR-HEDHSDRTIG 2540  
QY 117 QYGNNAALVNQTASDSSVM 136  
DB 2541 DLGGNGYSGTVGVRASSTL 2560  
RESULT 11  
ID Q8EIH3 PRELIMINARY; PRT; 139 AA.  
AC Q8EIH3;  
DT 01-MAR-2003 (TrEMBLrel. 23, Created)  
DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)  
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)  
DE Minor curlin subunit CsgB, putative.  
GN SO0866.  
OS Shewanella oneidensis.  
OC Bacteria; Proteobacteria; Gammaproteobacteria; Alteromonadales;  
OC Alteromonadaceae; Shewanella.  
OX NCBI\_TaxID=70863;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=MR-1;  
RX MEDLINE=22297686; PubMed=12368813;  
RA Heidelberg J.F., Paulsen I.T., Nelson K.E., Gaidos E.J., Nelson W.C.,  
RA Read T.D., Eisen J.A., Seshadri R., Ward N., Methe B., Clayton R.A.,  
RA Meyer T., Tsapin A., Scott J., Beanan M., Brinkac L., Daugherty S.,  
RA DeBoy R.T., Dodson R.J., Durkin A.S., Haft D.H., Kolonay J.F.,  
RA Madupu R., Peterson J.D., Umayam L.A., White O., Wolf A.M.,  
RA Vamathevan J., Weidman J., Ingham M., Lee K., Berry K., Lee C.,  
RA Mueller J., Khouri H., Gill J., Uitterlinden T.R., McDonald L.A.,  
RA Feldblyum T.V., Smith H.O., Venter J.C., Neilson K.H., Fraser C.M.;  
RT "Genome sequence of the dissimilatory metal ion-reducing bacterium Shewanella oneidensis."  
RL Nat. Biotechnol. 20:1118-1123(2002).  
DR EMBL; AE015532; AAN53942.1; --  
DR TIGR; SO0866; --  
KW Complete proteome.  
KW SEQUENCE 139 AA; 14811 MW; 41EC1CFA76957920 CRC64;  
Query Match 14.2%; Score 110; DB 16; Length 139;  
Best Local Similarity 28.3%; Pred. No. 0.13;  
Matches 30; Conservative 19; Mismatches 33; Indels 24; Gaps 3;  
QY 39 SGPDSTLSIYQVGSANAAALQSDARKSETTITQSGYNGADYDQLVTRVVTHEMAHAFR 98  
DB 41 SGDRNLIDVVOGTANQGVIFQSGSDNS-AVYTAGNDN-----ISLVTQGT----- 87  
QY 99 NNATIDQNAKNSDITVGQYGNNAALVNQTASDSSVMVQVGFN 144

Db 88 -----NNEVQLQVGAQKASITQIGNDNLVQLNGLSGN 122

RESULT 12

Q8EFU3 PRELIMINARY; PRT; 1422 AA.

AC Q8EFU3; 23, Created

DT 01-MAR-2003 (TEMBLrel. 23, Last sequence update)

DT 01-MAR-2003 (TEMBLrel. 23, Last sequence update)

DT 01-OCT-2003 (TEMBLrel. 25, Last annotation update)

DE Lipoprotein, putative.

GN S01872.

OS Shewanella oneidensis.

OC Bacteria; Proteobacteria; Gammaproteobacteria; Alteromonadales;

OC Alteromonadaceae; Shewanella.

OX NCBI\_TaxID=70863;

RN [1]

RP STRAIN=NR-1.

RC MEDLINE=2237686; PubMed=12368813;

RX Heidelberg J.F., Paulsen I.T., Nelson K.E., Gaidos E.J., Nelson W.C., Read T.D., Eisen J.A., Seshadri R., Ward N., Methe B., Clayton R.A., Meyer T., Tsapin A., Scott J., Beanan M., Brinkac L., Daugherty S., DeBoy R.T., Dodson R.J., Durkin A.S., Haft D.H., Kolonay J.F., Madupu R., Peterson J.D., Umayam L.A., White O., Wolf A.M., Vamathevan J., Weidman J., Imprim M., Lee K., Berry K., Lee C., Mueller J., Khouri H., Gill J., Uterback T.R., McDonald L.A., Feldblyum T.V., Smith H.O., Venter J.C., Nealon K.H., Fraser C.M.; Nat. Biotechnol. 20:1118-1123 (2002).

RL EMBL; AEO15630; AAN54924.1; --

DR TIGR; S01872; --

DR InterPro; IPR00437; Prok\_lipoprot.S.

DR PROSITE; PS00013; PROKAR\_LIPOPROTEIN; 1.

KW Complete proteome.

SQ SEQUENCE 1422 AA; 148253 MW; 840DB6300C993DB8 CRC64;

Query Match 13.9%; Score 108; DB 16; Length 1422;

Best Local Similarity 26.5%; Pred. No. 3.4; Indels 30; Gaps 5;

Matches 36; Conservative 23; Mismatches 47; Indels 30; Gaps 5;

Qy 7 AAFAAIIVGSGALAGVVPWGGGNGHNGSGDPSTLS----TYQYGSANAALALQSD 62

Db 1294 AAFAPKTDGS-----VVTWGGWGDGNGDSSAVALQTSVKAIYSTVSAFAAI----- 1343

Qy 63 ARKSETTITQSGYNGADYDQLVTRV-----THEMAHAFRNATIDQWNAKSDIT 114

Db 1344 -KDDSSVTVGGVWGGDSEVAGQTSVKAIYSTDSAFAAVKDGLVTVWGSNN--- 1399

Qy 115 VGOYGGNNAALNQTA 130

Db 1400 -----GGNSSSVASQLA 1411

RESULT 13

Q92U08 PRELIMINARY; PRT; 2174 AA.

AC Q92U08; 19, Created

DT 01-DEC-2001 (TEMBLrel. 19, Last sequence update)

DT 01-DEC-2001 (TEMBLrel. 19, Last sequence update)

DT 01-OCT-2003 (TEMBLrel. 25, Last annotation update)

DE Hypothetical glycine-rich protein SMB21548.

GN R30989 OR SMB21548.

OS Rhizobium meliloti (Sinorhizobium meliloti).

OC Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;

OC Rhizobiaceae; Sinorhizobium/Ensifer group; Sinorhizobium.

OX NCBI\_TaxID=382;

RN [1]

RP STRAIN=1021;

RP SEQUENCE FROM N.A.

Query Match 13.7%; Score 106.5; DB 16; Length 502;

Best Local Similarity 24.2%; Pred. No. 1.2; Indels 22; Gaps 61;

Matches 37; Conservative 21; Mismatches 61; Indels 22; Gaps 61;

Qy 11 AIVVGSALAGVVPQ--WGGGNGHNGSGDPSTLSIYQYGSANAALALQSDAR---K 65

Db 693 AITAGAGAVGIIAQSIGGGN--GGNATGGDAGFGSFGCGGGGGGYANTANVCFK 749

Qy 66 SETTITQSGYNGADYDQLVTRVVTHEMAHAFRNATIDQWNAK---NSDITVGOYGGNN 122

Db 750 GLTLTTQGGHAAG-----IVAQVGGCGGTGTGTASSYAGTGTASVAVGCTGGNG 800

Qy 123 AA--LVNQTSADSSVMVRQVG 141

Db 801 GAGGEVSVSLTDSAIRTGQGG 821

RESULT 14

Q8EIH4 PRELIMINARY; PRT; 502 AA.

AC Q8EIH4; 23, Created

DT 01-MAR-2003 (TEMBLrel. 23, Last sequence update)

DT 01-MAR-2003 (TEMBLrel. 23, Last sequence update)

DE Conserved hypothetical protein.

GN S00865.

OS Shewanella oneidensis.

OC Bacteria; Proteobacteria; Gammaproteobacteria; Alteromonadales;

OC Alteromonadaceae; Shewanella.

OX NCBI\_TaxID=70863;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=NR-1.

RX MEDLINE=2237686; PubMed=12368813;

RX Heidelberg J.F., Paulsen I.T., Nelson K.E., Gaidos E.J., Nelson W.C., Read T.D., Eisen J.A., Seshadri R., Ward N., Methe B., Clayton R.A., Meyer T., Tsapin A., Scott J., Beanan M., Brinkac L., Daugherty S., DeBoy R.T., Dodson R.J., Durkin A.S., Haft D.H., Kolonay J.F., Madupu R., Peterson J.D., Umayam L.A., White O., Wolf A.M., Vamathevan J., Weidman J., Imprim M., Lee K., Berry K., Lee C., Mueller J., Khouri H., Gill J., Uterback T.R., McDonald L.A., Feldblyum T.V., Smith H.O., Venter J.C., Nealon K.H., Fraser C.M.; Nat. Biotechnol. 20:1118-1123 (2002).

RL EMBL; AEO15532; AAN53941.1; --

DR TIGR; S00865; --

KW Hypothetical protein; Complete proteome.

SQ SEQUENCE 502 AA; 52441 MW; D08CA23D6C46B62D CRC64;

Query Match 13.7%; Score 106.5; DB 16; Length 502;

Best Local Similarity 24.2%; Pred. No. 1.2; Indels 22; Gaps 61;

Matches 37; Conservative 21; Mismatches 61; Indels 22; Gaps 61;

Qy 11 AIVVGSALAGVVPQ--WGGGNGHNGSGDPSTLSIYQYGSANAALALQSDAR---K 65

Db 693 AITAGAGAVGIIAQSIGGGN--GGNATGGDAGFGSFGCGGGGGGYANTANVCFK 749

Qy 66 SETTITQSGYNGADYDQLVTRVVTHEMAHAFRNATIDQWNAK---NSDITVGOYGGNN 122

Db 750 GLTLTTQGGHAAG-----IVAQVGGCGGTGTGTASSYAGTGTASVAVGCTGGNG 800

Qy 123 AA--LVNQTSADSSVMVRQVG 141

Db 801 GAGGEVSVSLTDSAIRTGQGG 821

RESULT 14

Q8EIH4 PRELIMINARY; PRT; 502 AA.

AC Q8EIH4; 23, Created

DT 01-MAR-2003 (TEMBLrel. 23, Last sequence update)

DT 01-MAR-2003 (TEMBLrel. 23, Last sequence update)

DE Conserved hypothetical protein.

GN S00865.

OS Shewanella oneidensis.

OC Bacteria; Proteobacteria; Gammaproteobacteria; Alteromonadales;

OC Alteromonadaceae; Shewanella.

OX NCBI\_TaxID=70863;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=NR-1.

RX MEDLINE=2237686; PubMed=12368813;

RX Heidelberg J.F., Paulsen I.T., Nelson K.E., Gaidos E.J., Nelson W.C., Read T.D., Eisen J.A., Seshadri R., Ward N., Methe B., Clayton R.A., Meyer T., Tsapin A., Scott J., Beanan M., Brinkac L., Daugherty S., DeBoy R.T., Dodson R.J., Durkin A.S., Haft D.H., Kolonay J.F., Madupu R., Peterson J.D., Umayam L.A., White O., Wolf A.M., Vamathevan J., Weidman J., Imprim M., Lee K., Berry K., Lee C., Mueller J., Khouri H., Gill J., Uterback T.R., McDonald L.A., Feldblyum T.V., Smith H.O., Venter J.C., Nealon K.H., Fraser C.M.; Nat. Biotechnol. 20:1118-1123 (2002).

RL EMBL; AEO15532; AAN53941.1; --

DR TIGR; S00865; --

KW Hypothetical protein; Complete proteome.

SQ SEQUENCE 502 AA; 52441 MW; D08CA23D6C46B62D CRC64;

Query Match 13.7%; Score 106.5; DB 16; Length 502;

Best Local Similarity 24.2%; Pred. No. 1.2; Indels 22; Gaps 61;

Matches 37; Conservative 21; Mismatches 61; Indels 22; Gaps 61;

Qy 11 AIVVGSALAGVVPQ--WGGGNGHNGSGDPSTLSIYQYGSANAALALQSDAR---K 65

Db 693 AITAGAGAVGIIAQSIGGGN--GGNATGGDAGFGSFGCGGGGGGYANTANVCFK 749

Qy 66 SETTITQSGYNGADYDQLVTRVVTHEMAHAFRNATIDQWNAK---NSDITVGOYGGNN 122

Db 750 GLTLTTQGGHAAG-----IVAQVGGCGGTGTGTASSYAGTGTASVAVGCTGGNG 800

Qy 123 AA--LVNQTSADSSVMVRQVG 141

Db 801 GAGGEVSVSLTDSAIRTGQGG 821

RESULT 14

Q8EIH4 PRELIMINARY; PRT; 502 AA.

AC Q8EIH4; 23, Created

DT 01-MAR-2003 (TEMBLrel. 23, Last sequence update)

DT 01-MAR-2003 (TEMBLrel. 23, Last sequence update)

DE Conserved hypothetical protein.

GN S00865.

OS Shewanella oneidensis.

OC Bacteria; Proteobacteria; Gammaproteobacteria; Alteromonadales;

OC Alteromonadaceae; Shewanella.

OX NCBI\_TaxID=70863;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=NR-1.

RX MEDLINE=2237686; PubMed=12368813;

RX Heidelberg J.F., Paulsen I.T., Nelson K.E., Gaidos E.J., Nelson W.C., Read T.D., Eisen J.A., Seshadri R., Ward N., Methe B., Clayton R.A., Meyer T., Tsapin A., Scott J., Beanan M., Brinkac L., Daugherty S., DeBoy R.T., Dodson R.J., Durkin A.S., Haft D.H., Kolonay J.F., Madupu R., Peterson J.D., Umayam L.A., White O., Wolf A.M., Vamathevan J., Weidman J., Imprim M., Lee K., Berry K., Lee C., Mueller J., Khouri H., Gill J., Uterback T.R., McDonald L.A., Feldblyum T.V., Smith H.O., Venter J.C., Nealon K.H., Fraser C.M.; Nat. Biotechnol. 20:1118-1123 (2002).

RL EMBL; AEO15532; AAN53941.1; --

DR TIGR; S00865; --

KW Hypothetical protein; Complete proteome.

SQ SEQUENCE 502 AA; 52441 MW; D08CA23D6C46B62D CRC64;

Query Match 13.7%; Score 106.5; DB 16; Length 502;

Best Local Similarity 24.2%; Pred. No. 1.2; Indels 22; Gaps 61;

Matches 37; Conservative 21; Mismatches 61; Indels



GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: August 2, 2004, 14:35:42 ; Search time 44.9 Seconds  
(without alignments)  
950.215 Million cell updates/sec

Title: US-09-543-407-30  
Perfect score: 768  
Sequence: 1 MKLLKVAFAAIVSGSALA.....DSSVMVRGVGNATATQY 151

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 1586107 seqs, 282547505 residues

Total number of hits satisfying chosen parameters: 1586107

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : A Geneseq 29Jan04: \*  
1: geneseqp1980s: \*  
2: geneseqp1990s: \*  
3: geneseqp2000s: \*  
4: geneseqp2001s: \*  
5: geneseqp2002s: \*  
6: geneseqp2003as: \*  
7: geneseqp2003bs: \*  
8: geneseqp2004s: \*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	768	100.0	151	3	AAB36355 Agfa::PT3
2	700	91.1	151	3	AAB36352
3	681	88.7	151	2	Aar74625 Agfa sequ
4	681	88.7	151	3	AAB36341 Salmonell
5	676	88.0	151	2	Aaw23570 Salmonell
6	659	85.8	151	3	AAB36347 Agfa::PT3
7	657	85.5	151	3	AAB36351 Agfa::PT3
8	613	79.8	151	3	AAB36334 Agfa::PT3
9	603	78.5	151	3	AAB36346 Agfa::PT3
10	600	78.1	151	3	AAB36333 Agfa::PT3
11	599	78.0	151	3	AAB36349 Agfa::PT3
12	597	77.7	151	3	AAB36350 Agfa::PT3
13	566	73.7	151	3	AAB36348 Agfa::PT3
14	517	67.3	151	3	AAB36343 Escherich
15	512	66.7	151	7	Abr82651 E. coli C
16	496	64.6	120	2	Aar62761 Agfa sequ
17	496	64.6	120	2	Aaw23569 Salmonell
18	439	57.2	142	2	AAR52664 Fibronect
19	367	47.8	122	2	AAR52663 FNB curli
20	215	28.0	45	3	AAB36316 Salmonell
21	132	17.2	22	3	AAB36318 Salmonell
22	115	15.0	22	3	AAB36325 Salmonell
23	115	15.0	22	3	AAB36339 Salmonell
24	115	15.0	22	3	AAB36320 Salmonell
25	113	14.7	24	7	Abr82644 E. coli c

26 109 14.2 23 3 AAB36340 Salmonell  
27 109 14.2 23 3 AAB36324 Salmonell  
28 109 14.2 23 3 AAB36319 Salmonell  
29 104.5 13.6 151 3 AAB36344 Escherich  
30 102 13.3 26 7 ABR82649 E. coli V  
31 101.5 13.2 520 6 AAO16497 Argiope t  
32 99.5 13.0 151 3 AAB36342 Salmonell  
33 98 12.8 26 7 ABR82645 E. coli C  
34 96 12.5 19 3 AAB36323 Salmonell  
35 96 12.5 19 3 AAB36336 Salmonell  
36 96 12.5 19 3 AAB36328 Salmonell  
37 95.5 12.4 908 4 ABB65007 Drosophil  
38 95 12.4 718 3 AAY78364 H. pylori  
39 94.5 12.3 908 4 ABB62355 Drosophil  
40 94 12.2 354 7 ABO23520 Mycobacte  
41 93.5 12.2 2204 6 ABU36440 Protein e  
42 92.5 12.0 738 2 AAU56163 New DNA s  
43 92.5 12.0 850 4 ABB65764 Drosophil  
44 92.5 12.0 1028 4 ABB62708 Drosophil  
45 92.5 12.0 1721 6 ABU34624 Protein e

## ALIGNMENTS

RESULT 1  
AAB36355  
ID AAB36355 standard; protein; 151 AA.  
XX  
AC AAB36355;  
XX  
DT 26-FEB-2001 (first entry)  
XX  
DE Agfa::PT3#10 amino acid sequence SEQ ID NO:30.  
XX  
KW Salmonella; agfa; chromosomal gene replacement; fimbrin; epitope;  
KW vaccine; immune response; immunogen.  
XX  
OS Salmonella enteritidis.  
OS Escherichia coli.  
OS Synthetic.  
XX  
PN WO200060102-A2.  
XX  
PD 12-OCT-2000.  
XX  
PF 05-APR-2000; 2000WO-CA000356.  
XX  
PR 05-APR-1999; 99US-0127888P.  
XX (UUVI-) UNIV VICTORIA.  
XX White AP, Doran JL, Collison SK, Kay WW;  
XX WPI; 2000-672631/65.  
XX N-PSDB; AAC64631.  
XX  
XX Recombinant agfa gene having a segment replaced by a foreign DNA sequence  
PT which encodes foreign epitope or antigen, expresses recombinant Agfa  
PT protein useful for eliciting immune response in animal.  
XX  
PS Disclosure; Page 139; 139pp; English.  
XX  
XX The present invention describes a recombinant agfa gene (I) where a  
CC segment of the gene has been replaced by a segment of a foreign DNA  
CC sequence which encodes a foreign epitope or antigen. Also described are:  
CC (1) use of thin aggregative fimbriae (SF17/TAF) nucleation depended  
CC assembly system of strains of Salmonella, Escherichia coli and  
CC Enterobacteriaceae for the production of fimbriae comprising recombinant  
CC Agfa, CsgA and Agfa-homologue fimbria subunits, respectively; (2)  
CC directing recombination of a recombinant gene into the chromosome of the  
CC homologous species; (3) directing recombination of a recombinant gene  
CC back into the chromosome of the homologous species, replacing the native

CC copy of that gene; and (4) eliciting an immune response in an animal,  
 CC comprising separating an amino acid polymer comprising a recombinant Agfa  
 CC protein containing a replacement segment or segments of a foreign amino  
 CC acid sequence or sequences grown on a Salmonella, E. coli or  
 CC Enterobacteriaceae host cell, from the host cell and introducing the  
 CC polymer into the animal in conjunction with a carrier or diluent. (1) is  
 CC useful for the expression of recombinant Agfa protein which is useful for  
 CC eliciting an immune response in an animal. In a fimbrial presentation  
 CC system the heterologous antigens are presented in high numbers (up to  
 CC 500,000 copies/cell), the hybrid fimbrial protein possesses both the  
 CC immunogenicity and adhesion properties relevant for an efficient live  
 CC vaccine, the carrier fimbrial subunit proteins are usually strong  
 CC immunogens, which may be important for directing an immune response  
 CC against the inserted epitope, and hybrid fimbriae are easy and  
 CC inexpensive to purify in large amount. The present sequence is given in  
 CC the exemplification of the present invention  
 CC  
 CC Sequence 151 AA;  
 CC  
 CC

Query Match 100.0%; Score 768; DB 3; Length 151;  
 Best Local Similarity 100.0%; Pred. No. 1.7e-66;  
 Matches 151; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 M K L L K V A A F A A I V V G S A L A G V V P Q W G G G N H N G G N S G P D S T L S I Y Q Y G S A N A A L A Q 60  
 D B 1 M K L L K V A A F A A I V V G S A L A G V V P Q W G G G N H N G G N S G P D S T L S I Y Q Y G S A N A A L A Q 60  
 QY 61 S D A R K S E T T I T O S G Y G N G A D V G G A D N S T I E L T Q N F R N N A T Y D Q L V T R V V T H E M A H A G G 120  
 D B 61 S D A R K S E T T I T O S G Y G N G A D V G G A D N S T I E L T Q N F R N N A T Y D Q L V T R V V T H E M A H A G G 120  
 QY 121 N N A A L V N Q T A S D S S V M V R Q V G F G N N A T A N Q Y 151  
 D B 121 N N A A L V N Q T A S D S S V M V R Q V G F G N N A T A N Q Y 151

RESULT 2  
 ID AAB36352 standard; protein; 151 AA.  
 AC AAB36352;  
 XX 26-FEB-2001 (first entry)  
 DE Agfa::PT3#7 amino acid sequence SEQ ID NO:24.  
 XX Salmonella; agfa; chromosomal gene replacement; fimbrial; epitope;  
 KW vaccine; immune response; immunogen.  
 XX Salmonella enteritidis.  
 OS Escherichia coli.  
 OS Synthetic.  
 XX WO200006102-A2.  
 XX 12-OCT-2000.  
 XX 05-APR-2000; 2000WO-CA000356.  
 XX 05-APR-1999; 99US-0127888P.  
 XX (UYVI-) UNIV VICTORIA.  
 XX White AP, Doran JL, Collison SK, Kay WW,  
 XX WPI; 2000-672631/65.  
 DR N-PSDB; AAC64628.  
 XX Recombinant agfa gene having a segment replaced by a foreign DNA sequence  
 PT which encodes foreign epitope or antigen, expresses recombinant Agfa  
 PT protein useful for eliciting immune response in animal.  
 XX  
 PS Disclosure; Page 138; 139pp; English.

XX The present invention describes a recombinant agfa gene (1) where a  
 CC segment of the gene has been replaced by a segment of a foreign DNA  
 CC sequence which encodes a foreign epitope or antigen. Also described are:  
 CC (1) use of thin aggregative fimbriae (SEF17/TAF) nucleation depended  
 CC assembly system of strains of Salmonella, Escherichia coli and  
 CC Enterobacteriaceae for the production of fimbriae comprising recombinant  
 CC Agfa, CsgA and Agfa-homologue fimbrial subunits, respectively; (2)  
 CC directing recombination of a recombinant gene into the chromosome of the  
 CC homologous species; (3) directing recombination of a recombinant gene  
 CC back into the chromosome of the homologous species, replacing the native  
 CC copy of that gene; and (4) eliciting an immune response in an animal,  
 CC comprising separating an amino acid polymer comprising a recombinant Agfa  
 CC protein containing a replacement segment or segments of a foreign amino  
 CC acid sequence or sequences grown on a Salmonella, E. coli or  
 CC Enterobacteriaceae host cell, from the host cell and introducing the  
 CC polymer into the animal in conjunction with a carrier or diluent. (1) is  
 CC useful for the expression of recombinant Agfa protein which is useful for  
 CC eliciting an immune response in an animal. In a fimbrial presentation  
 CC system the heterologous antigens are presented in high numbers (up to  
 CC 500,000 copies/cell), the hybrid fimbrial protein possesses both the  
 CC immunogenicity and adhesion properties relevant for an efficient live  
 CC vaccine, the carrier fimbrial subunit proteins are usually strong  
 CC immunogens, which may be important for directing an immune response  
 CC against the inserted epitope, and hybrid fimbriae are easy and  
 CC inexpensive to purify in large amount. The present sequence is given in  
 CC the exemplification of the present invention  
 CC  
 CC Sequence 151 AA;  
 CC  
 CC

Query Match 91.1%; Score 700; DB 3; Length 151;  
 Best Local Similarity 89.9%; Pred. No. 6.8e-60;  
 Matches 143; Conservative 0; Mismatches 0; Indels 16; Gaps 2;  
 QY 1 M K L L K V A A F A A I V V G S A L A G V V P Q W G G G N H N G G N S G P D S T L S I Y Q Y G S A N A A L A Q 60  
 D B 1 M K L L K V A A F A A I V V G S A L A G V V P Q W G G G N H N G G N S G P D S T L S I Y Q Y G S A N A A L A Q 60  
 QY 61 S D A R K S E T T I T O S G Y G N G A D V G G A D N S T I E L T Q N F R N N A T Y D Q L V T R V V T 112  
 D B 61 S D A R K S E T T I T O S G Y G N G A D V G G A D N S T I E L T Q N F R N N A T Y D Q L V T R V V T 120  
 QY 113 H E M A H A G G N N A A L V N Q T A S D S S V M V R Q V G F G N N A T A N Q Y 151  
 D B 121 H E M A H A -----N Q T A S D S S V M V R Q V G F G N N A T A N Q Y 151

RESULT 3  
 ID AAR74625 standard; protein; 151 AA.  
 AC AAR74625;  
 XX 25-MAR-2003 (revised)  
 DT 26-JUN-1995 (first entry)  
 DE Agfa sequence.  
 XX Salmonella; Agfa; vaccine.  
 XX Salmonella.  
 OS WO9425598-A2.  
 XX 10-NOV-1994.  
 XX 26-APR-1994; 94WO-IB000207.  
 XX 26-APR-1993; 93US-00054452.  
 XX (UYVI-) UNIV VICTORIA INNOVATION & DEV CORP.  
 XX (KING/) KING J.

PI -Kay WW, Collinson SK, Clouthier SC, Doran JL;  
XX WPI; 1994-358275/44.  
DR N-PSDB; AAQ87467.  
XX  
XX  
PT Eliciting an immune response to Salmonella - using attenuated Salmonella  
PT strains, vector constructs, or compsns. contg. fimbrial type proteins.  
XX  
XX Disclosure; Fig 7B; 95pp; English.  
XX  
XX The Salmonella Agfa protein and DNA are used in vaccine and genetic  
CC immunization compositions, respectively, to elicit an immune response to  
CC Salmonella in animals (e.g. food producing animals) and humans. (Updated  
CC on 25-MAR-2003 to correct PN field.)  
XX  
SQ Sequence 151 AA;  
Query Match 88.7%; Score 681; DB 2; Length 151;  
Best Local Similarity 90.7%; Pred. No. 4.7e-58;  
Matches 137; Conservative 2; Mismatches 12; Indels 0; Gaps 0;  
QY 1 MKLLKVAFAAIVVSGSALAGVVPQWGGGNGNNGSGPDSLTSLIYQYGSANAALALQ 60  
DB 1 MKLLKVAFAAIVVSGSALAGVVPQWGGGNGNNGSGPDSLTSLIYQYGSANAALALQ 60  
QY 61 SDARKSETTITQSGYNGADVGQGDNSTIETONGFRNNATYDOLVTRVVTHEMAHAGG 120  
DB 61 SDARKSETTITQSGYNGADVGQGDNSTIETONGFRNNATIDQWNAKNSDITVQYGG 120  
QY 121 NNAALVNQTSASDSSVMVRQVGFNNATANQY 151  
DB 121 NNAALVNQTSASDSSVMVRQVGFNNATANQY 151  
RESULT 4  
AAB36341 ID AAB36341 standard; protein; 151 AA.  
XX  
XX AAB36341;  
XX  
DT 26-FEB-2001 (first entry)  
DE  
DE Salmonella enteritidis Agfa amino acid sequence SEQ ID NO:5.  
XX  
XX Salmonella; agfa; chromosomal gene replacement; fimbria; epitope;  
KW vaccine; immune response; immunogen.  
XX  
XX Salmonella enteritidis.  
OS  
XX WO200060102-A2.  
FN  
XX  
XX 12-OCT-2000.  
PD  
XX 05-APR-2000; 2000WO-CA000356.  
PF  
XX 05-APR-1999; 99US-0127888P.  
PR  
XX (UYVI-) UNIV VICTORIA.  
PA  
XX White AP, Doran JL, Collinson SK, Kay WW;  
PI  
XX WPI; 2000-672631/65.  
DR  
XX N-PSDB; AAC64617.  
DR  
XX Recombinant agfa gene having a segment replaced by a foreign DNA sequence  
PT which encodes foreign epitope or antigen, expresses recombinant Agfa  
PT protein useful for eliciting immune response in animal.  
XX  
XX Disclosure; Page 135; 139pp; English.  
XX  
XX The present invention describes a recombinant agfa gene (1) where a  
CC segment of the gene has been replaced by a segment of a foreign DNA  
CC sequence which encodes a foreign epitope or antigen. Also described are:

CC (1) use of thin aggregative fimbriae (SEF17/TAF) nucleation depended  
CC assembly system of strains of Salmonella, Escherichia coli and  
CC Enterobacteriaceae for the production of fimbriae comprising recombinant  
CC Agfa, CsgA and Agfa-homologue fimbrial subunits, respectively; (2)  
CC directing recombination of a recombinant gene into the chromosome of the  
CC homologous species; (3) directing recombination of a recombinant gene  
CC back into the chromosome of the homologous species, replacing the native  
CC copy of that gene; and (4) eliciting an immune response in an animal,  
CC comprising separating an amino acid polymer comprising a recombinant Agfa  
CC protein containing a replacement segment or segments of foreign amino  
CC acid sequence or sequences grown on a Salmonella, E. coli or  
CC Enterobacteriaceae host cell, from the host cell and introducing the  
CC polymer into the animal in conjunction with a carrier or diluent. (1) is  
CC useful for the expression of recombinant Agfa protein which is useful for  
CC eliciting an immune response in an animal. In a fimbrial presentation  
CC system the heterologous antigens are presented in high numbers (up to  
CC 500,000 copies/cell), the hybrid fimbrial protein possesses both the  
CC immunogenicity and adhesion properties relevant for an efficient live  
CC vaccine, the carrier fimbrial subunit proteins are usually strong  
CC immunogens, which may be important for directing an immune response  
CC against the inserted epitope, and hybrid fimbriae are easy and  
CC inexpensive to purify in large amount. The present sequence is given in  
CC the exemplification of the present invention  
XX  
SQ Sequence 151 AA;  
Query Match 88.7%; Score 681; DB 3; Length 151;  
Best Local Similarity 90.7%; Pred. No. 4.7e-58;  
Matches 137; Conservative 2; Mismatches 12; Indels 0; Gaps 0;  
QY 1 MKLLKVAFAAIVVSGSALAGVVPQWGGGNGNNGSGPDSLTSLIYQYGSANAALALQ 60  
DB 1 MKLLKVAFAAIVVSGSALAGVVPQWGGGNGNNGSGPDSLTSLIYQYGSANAALALQ 60  
QY 61 SDARKSETTITQSGYNGADVGQGDNSTIETONGFRNNATYDOLVTRVVTHEMAHAGG 120  
DB 61 SDARKSETTITQSGYNGADVGQGDNSTIETONGFRNNATIDQWNAKNSDITVQYGG 120  
QY 121 NNAALVNQTSASDSSVMVRQVGFNNATANQY 151  
DB 121 NNAALVNQTSASDSSVMVRQVGFNNATANQY 151  
RESULT 5  
AAW23570 ID AAW23570 standard; protein; 151 AA.  
XX  
XX AAW23570;  
XX  
XX 25-MAR-2003 (revised)  
DT 29-SEP-1997 (first entry)  
XX  
XX Salmonella enteritidis 27655-3b agfa.  
DE  
XX Enteropathogenic bacteria; enterobacteria; S.enteritidis; antibody.  
KW  
XX Salmonella enteritidis.  
OS  
XX Key Location/Qualifiers  
FH Misc-difference 123  
FT /note= "Encoded by GCC"  
FN  
XX US5635617-A.  
XX  
XX 03-JUN-1997.  
PD  
XX 26-APR-1994; 94US-00233788.  
PF  
XX 26-APR-1993; 93US-00054452.  
PR  
XX (UYVI-) UNIV VICTORIA INNOVATION & DEV CORP.  
PA  
XX Collinson SK, Kay WW, Doran JL;  
PI

XX WPI: 1997-309886/28.  
DR N-PSDB; AAR74342.  
XX  
XX Isolated Salmonella gene agfa - used for diagnosis of Salmonella or  
PT enteropathogenic bacteria of the Enterobacteriaceae family.  
XX  
XX Example 2; Fig 7; 85pp; English.  
XX  
XX The present sequence represents agfa encoded by the full agfa gene  
CC derived from Salmonella enteritidis 27653-3b. The nucleic acid can be  
CC used to provide diagnostic assays for Salmonella and/or enteropathogenic  
CC bacteria of the family Enterobacteriaceae. It can also be used to provide  
CC proteins and antibodies which can be used for assays. The nucleic acid  
CC sequence can be used to provide probes or primers which can specifically  
CC hybridise to nucleic acid molecules from greater than 99% of Salmonella  
CC strains that are pathogenic to warm-blooded animals relative to nucleic  
CC acid molecules from virtually all other microbial organisms. (Updated on  
CC 25-MAR-2003 to correct PF field.)  
XX  
XX Sequence 151 AA;  
SQ  
Query Match 88.0%; Score 676; DB 2; Length 151;  
Best Local Similarity 90.1%; Pred. No. 1.4e-57;  
Matches 136; Conservative 2; Mismatches 13; Indels 0; Gaps 0;  
QY 1 MKLLKVAFAAIVVSGSALAGVVPQWGGGNGGNGSGPDSLTLSIYQGSANAALALQ 60  
DB 1 MKLLKVAFAAIVVSGSALAGVVPQWGGGNGGNGSGPDSLTLSIYQGSANAALALQ 60  
QY 61 SDARKSETTITQSGYNGADVGQADNSTIELTQNGFRNNATYDQLVTVVTHEMAHAGG 120  
DB 61 SDARKSETTITQSGYNGADVGQADNSTIELTQNGFRNNATYDQLVTVVTHEMAHAGG 120  
QY 121 NNAALVNTQASDSSVMVRQVGFNNATANQY 151  
DB 121 NNPALVNTQASDSSVMVRQVGFNNATANQY 151  
RESULT 6  
AAB36347  
ID AAB36347 standard; protein; 151 AA.  
XX  
XX AAB36347;  
XX  
XX 26-FEB-2001 (first entry)  
XX  
XX Agfa::PT3#2 amino acid sequence SEQ ID NO:14.  
XX  
XX Salmonella; agfa; chromosomal gene replacement; fimbrin; epitope;  
KW vaccine; immune response; immunogen.  
XX  
XX Salmonella enteritidis.  
OS Escherichia coli.  
OS Synthetic.  
XX  
XX W0200060102-A2.  
XX  
XX 12-OCT-2000.  
XX  
XX 05-APR-2000; 2000WO-CA000356.  
XX  
XX 05-APR-1999; 99US-0127888P.  
XX  
XX (UUVI-) UNIV VICTORIA.  
XX  
XX White AP, Doran JL, Collison SK, Kay WH;  
XX  
XX WPI: 2000-672631/65.  
DR N-PSDB; AAC64623.  
XX  
XX Recombinant agfa gene having a segment replaced by a foreign DNA sequence  
PT which encodes foreign epitope or antigen, expresses recombinant Agfa

PT protein useful for eliciting immune response in animal.  
XX  
XX Disclosure; Page 136; 139pp; English.  
XX  
XX The present invention describes a recombinant agfa gene (I) where a  
CC segment of the gene has been replaced by a segment of a foreign DNA  
CC sequence which encodes a foreign epitope or antigen. Also described are:  
CC (1) use of thin aggregative fimbriae (SEF17/7AF) nucleation depended  
CC assembly system of strains of Salmonella, Escherichia coli and  
CC Enterobacteriaceae for the production of fimbriae comprising recombinant  
CC Agfa, CsgA and Agfa-homologue fimbrial subunits, respectively; (2)  
CC directing recombination of a recombinant gene into the chromosome of the  
CC homologous species; (3) directing recombination of a recombinant gene  
CC back into the chromosome of the homologous species, replacing the native  
CC copy of that gene; and (4) eliciting an immune response in an animal,  
CC comprising separating an amino acid polymer comprising a recombinant Agfa  
CC protein containing a replacement segment or segments of foreign amino  
CC acid sequence or sequences grown on a Salmonella, E. coli or  
CC Enterobacteriaceae host cell, from the host cell and introducing the  
CC polymer into the animal in conjunction with a carrier or diluent. (I) is  
CC useful for the expression of recombinant Agfa protein which is useful for  
CC eliciting an immune response in an animal. In a fimbrial presentation  
CC system the heterologous antigens are presented in high numbers (up to  
CC 500,000 copies/cell), the hybrid fimbrial protein possesses both the  
CC immunogenicity and adhesion properties relevant for an efficient live  
CC vaccine, the carrier fimbrial subunit proteins are usually strong  
CC immunogens, which may be important for directing an immune response  
CC against the inserted epitope, and hybrid fimbriae are easy and  
CC inexpensive to purify in large amount. The present sequence is given in  
CC the exemplification of the present invention  
XX  
XX Sequence 151 AA;  
SQ  
Query Match 85.8%; Score 659; DB 3; Length 151;  
Best Local Similarity 81.9%; Pred. No. 6.4e-56;  
Matches 136; Conservative 0; Mismatches 0; Indels 30; Gaps 2;  
QY 1 MKLLKVAFAAIVVSGSALAGVVPQWGGGNGGNGSGPDSLTLSIYQGSANAALALQ 60  
DB 1 MKLLKVAFAAIVVSGSALAGVVPQWGGGNGGNGSGPDSLTLSIYQGSANAALALQ 60  
QY 61 SDARKSETTITQSGYNGADVGQADNSTIELTQNGFRNNATYDQLVTVVTHEMAHAGG 105  
DB 61 SDARKSETTITQSGYNGADVGQADNSTIELTQNGFRNNATYDQLVTVVTHEMAHAGG 105  
QY 106 LVTRVTVTHEMAHAGGNAALVNTQASDSSVMVRQVGFNNATANQY 151  
DB 121 LVTRVTVTHEMAHA-----SVNVRQVGFNNATANQY 151  
RESULT 7  
AAB36351  
ID AAB36351 standard; protein; 151 AA.  
XX  
XX AAB36351;  
XX  
XX 26-FEB-2001 (first entry)  
XX  
XX Agfa::PT3#6 amino acid sequence SEQ ID NO:22.  
XX  
XX Salmonella; agfa; chromosomal gene replacement; fimbrin; epitope;  
KW vaccine; immune response; immunogen.  
XX  
XX Salmonella enteritidis.  
OS Escherichia coli.  
OS Synthetic.  
XX  
XX W0200060102-A2.  
XX  
XX 12-OCT-2000.  
XX  
XX 05-APR-2000; 2000WO-CA000356.  
XX



```

PR 05-APR-1999; 99US-0127888P.
XX (UYVI-) UNIV VICTORIA.
XX
XX White AP, Doran JL, Collison SK, Kay WW;
XX
XX WPI; 2000-672631/65.
XX N-PSDB; AAC64627.
XX
XX Recombinant agfA gene having a segment replaced by a foreign DNA sequence
XX which encodes foreign epitope or antigen, expresses recombinant AgfA
XX protein useful for eliciting immune response in animal.
XX
XX Disclosure; Page 137; 139pp; English.
XX
XX The present invention describes a recombinant agfA gene (1) where a
XX segment of the gene has been replaced by a segment of a foreign DNA
XX sequence which encodes a foreign epitope or antigen. Also described are:
XX (1) use of thin aggregative fimbriae (SEF17/TAf) nucleation depended
XX assembly system of strains of Salmonella, Escherichia coli and
XX Enterobacteriaceae for the production of fimbriae comprising recombinant
XX AgfA, CsgA and agfA-homologue fimbria subunits, respectively; (2)
XX directing recombination of a recombinant gene into the chromosome of the
XX homologous species; (3) directing recombination of a recombinant gene
XX back into the chromosome of the homologous species, replacing the native
XX copy of that gene; and (4) eliciting an immune response in an animal,
XX comprising separating an amino acid polymer comprising a recombinant AgfA
XX protein containing a replacement segment or segments of foreign amino
XX acid sequence or sequences grown on a Salmonella, E. coli or
XX Enterobacteriaceae host cell, from the host cell and introducing the
XX polymer into the animal in conjunction with a carrier or diluent. (1) is
XX useful for the expression of recombinant AgfA protein which is useful for
XX eliciting an immune response in an animal. In a fimbrial presentation
XX system the heterologous antigens are presented in high numbers (up to
XX 500,000 copies/cell), the hybrid fimbria protein possesses both the
XX immunogenicity and adhesion properties relevant for an efficient live
XX vaccine, the carrier fimbrial subunit proteins are usually strong
XX immunogens, which may be important for directing an immune response
XX against the inserted epitope, and hybrid fimbriae are easy and
XX inexpensive to purify in large amount. The present sequence is given in
XX the exemplification of the present invention
XX
XX Sequence 151 AA;
XX
XX Query Match 85.5%; Score 657; DB 3; Length 151;
XX Best Local Similarity 81.9%; Pred. No.1e-55;
XX Matches 136; Conservative 0; Mismatches 0; Indels 30; Gaps 2
XX
XX QY 1 MKLLKVAFAAIVSGSALAGVWPQWGGGNGHNGCGNSGPDSTLSIYQGSAAALALQ 60
XX DB 1 MKLLKVAFAAIVSGSALAGVWPQWGGGNGHNGCGNSGPDSTLSIYQGSAAALALQ 60
XX
XX QY 61 SDARKSETITQSGYGNAGDVQGGADNSTIELTQGFPRNNATYDQLVTRVVTTHMAHA-- 118
XX DB 61 SDARKSETITQSGYGNAGDVQGGADN-----YDQLVTRVVTTHMAHADQ 105
XX
XX QY 119 -----GNNAAALVNOTASDSSVMVQVFGNNATANQY 151
XX DB 106 WNAKNSDITVGYQGGNAAALVNOTASDSSVMVQVFGNNATANQY 151
XX
XX RESULT 8
XX AAB36354
XX ID AAB36354 standard; protein; 151 AA.
XX AC
XX AAB36354;
XX AC
XX XX
XX 26-FEB-2001 (first entry)
XX
XX AgfA::P73#9 amino acid sequence SEQ ID NO:28.
XX
XX Salmonella; agfA; chromosomal gene replacement; fimbria; epitope;
XX vaccine; immune response; immunogen.
XX

```

	XX	Salmonella enteritidis.
	OS	Escherichia coli.
	OS	Synthetic.
	XX	WO200606102-A2.
	XX	12-OCT-2000.
	PD	
	PP	05-APR-2000; 2000WO-CA000356.
	PF	
	XX	05-APR-1999; 99US-0127888P.
	PR	(UYVI-) UNIV VICTORIA.
	XX	
	PA	
	XX	
	PI	White AP, Doran JL, Collison SK, Kay WW;
	XX	WPI; 2000-672631/65.
	DR	N-PSDB; AAC64630.
	DR	
	XX	
	PT	Recombinant agfa gene having a segment replaced by a foreign DNA sequence
	PT	which encodes foreign epitope or antigen, expresses recombinant Agfa
	PT	protein useful for eliciting immune response in animal.
	PT	
	PP	Disclosure; Page 138; 139pp; English.
	XX	
	CC	The present invention describes a recombinant agfa gene (1) where a
	CC	segment of the gene has been replaced by a segment of a foreign DNA
	CC	sequence which encodes a foreign epitope or antigen. Also described are:
	CC	(1) use of thin aggregative fimbriae (SEF17/TAF) nucleation depended
	CC	assembly system of strains of Salmonella, Escherichia coli and
	CC	Enterobacteriaceae for the production of fimbriae comprising recombinant
	CC	Agfa, CsgA and Agfa-homologue fimbrial subunits, respectively; (2)
	CC	directing recombination of a recombinant gene into the chromosome of the
	CC	homologous species; (3) directing recombination of a recombinant gene
	CC	back into the chromosome of the homologous species, replacing the native
	CC	copy of that gene; and (4) eliciting an immune response in an animal,
	CC	comprising separating an amino acid polymer comprising a recombinant Agfa
	CC	protein containing a replacement segment or segments of foreign amino
	CC	acid sequence or sequences grown on a Salmonella, E. coli or
	CC	Enterobacteriaceae host cell, from the host cell and introducing the
	CC	polymer into the animal in conjunction with a carrier or diluent. (1) is
	CC	useful for the expression of recombinant Agfa protein which is useful for
	CC	eliciting an immune response in an animal. In a fimbrial presentation
	CC	system the heterologous antigens are presented in high numbers (up to
	CC	500,000 copies/cell), the hybrid fimbrial protein possesses both the
	CC	immunogenicity and adhesion properties relevant for an efficient live
	CC	vaccine, the carrier fimbrial subunit proteins are usually strong
	CC	immunoogens, which may be important for directing an immune response
	CC	against the inserted epitope, and hybrid fimbriae are easy and
	CC	inexpensive to purify in large amount. The present sequence is given in
	CC	the exemplification of the present invention
	XX	
	XX	Sequence 151 AA;
	QQ	Query Match 79.8%; Score 613; DB 3; Length 151;
	XX	Best Local Similarity 74.6%; Pred. No. 1.9e-51;
	XX	Matches 129; Conservative 0; Mismatches 0; Indels 44; Gaps 2
	QY	1 MKLLKVAFAAIIVSGSALAGVWPQWGCGGNHNGGNSGSPDSTLSIYYGGSNNAALAQ 60
	Db	1 MKLLKVAFAAIIVSGSALAGVWPQWGCGGNHNGGNSGSPDSTLSIYYGGSNNAALAQ 60
	QY	61 SDARKSETITTSQGVNGADVGQCADNSTIELTQGNFRNNATVDQLVTRVTWTHEMAHA-- 118
	Db	61 SDARKSETITTSQGVNGAD-----YDQLVTRVTWTHEMAHFR 98
	QY	119 -----GGNNAALVNQTASDSSVMVRQVGFGNNTANQY 151
	Db	99 NNATIDOWNAKNSDITVGOYGNNAAALVNQTASDSSVMVRQVGFGNNTANQY 151

## RESULT 9

AAB36346  
ID AAB36346 standard; protein; 151 AA.  
XX  
AC AAB36346;  
XX  
DT 26-FEB-2001 (first entry)  
XX  
DE Agfa::PT3#1 amino acid sequence SEQ ID NO:12.  
XX  
KW Salmonella; agfa; chromosomal gene replacement; fimbrin; epitope;  
XX vaccine; immune response; immunogen.  
XX  
OS Salmonella enteritidis.  
OS Escherichia coli.  
OS Synthetic.  
XX  
FN WO200060102-A2.  
XX  
PD 12-OCT-2000.  
XX  
PF 05-APR-2000; 2000WO-CA000356.  
XX  
PR 05-APR-1999; 99US-0127888P.  
XX  
PA (UYVI-) UNIV VICTORIA.  
XX  
PI White AP, Doran JL, Collison SK, Kay WW;  
XX  
DR WPI; 2000-672631/65.  
DR N-PSDB; AAC64622.  
XX  
PT Recombinant agfa gene having a segment replaced by a foreign DNA sequence  
PT which encodes foreign epitope or antigen, expresses recombinant Agfa  
PT protein useful for eliciting immune response in animal.  
XX  
PS Disclosure; Page 135; 139pp; English.  
XX  
CC The present invention describes a recombinant agfa gene (I) where a  
CC segment of the gene has been replaced by a segment of a foreign DNA  
CC sequence which encodes a foreign epitope or antigen. Also described are:  
CC (1) use of thin aggregative fimbriae (SEF17/TAF) nucleation depended  
CC assembly system of strains of Salmonella, Escherichia coli and  
CC Enterobacteriaceae for the production of fimbriae comprising recombinant  
CC Agfa, CsgA and Agfa-homologue fimbria subunits, respectively; (2)  
CC directing recombination of a recombinant gene into the chromosome of the  
CC homologous species; (3) directing recombination of a recombinant gene  
CC back into the chromosome of the homologous species, replacing the native  
CC copy of that gene; and (4) eliciting an immune response in an animal,  
CC comprising separating an amino acid polymer comprising a recombinant Agfa  
CC protein containing a replacement segment or segments of foreign amino  
CC acid sequence or sequences grown on a Salmonella, E. coli or  
CC Enterobacteriaceae host cell, from the host cell and introducing the  
CC polymer into the animal in conjunction with a carrier or diluent. (I) is  
CC useful for the expression of recombinant Agfa protein which is useful for  
CC eliciting an immune response in an animal. In a fimbrial presentation  
CC system the heterologous antigens are presented in high numbers (up to  
CC 500,000 copies/cell), the hybrid fimbria protein possesses both the  
CC immunogenicity and adhesion properties relevant for an efficient live  
CC vaccine, the carrier fimbrial subunit proteins are usually strong  
CC immunogens, which may be important for directing an immune response  
CC against the inserted epitope, and hybrid fimbriae are easy and  
CC inexpensive to purify in large amount. The present sequence is given in  
CC the exemplification of the present invention  
XX  
SQ Sequence 151 AA;  
Query Match 78.5%; Score 603; DB 3; Length 151;  
Best Local Similarity 80.8%; Pred. No. 1.7e-50;  
Matches 122; Conservative 6; Mismatches 23; Indels 0; Gaps 0;  
1 MKLLKVAFAAIVVSGSALAGVVPQWGGGNGHNGGNSGPDSTLSIYQGSANAALALQ 60  
1 MKLLKVAFAAIVVSGSALAGVVPQWGGGNGHNGGNSGPDSTLSIYQGSANAALALQ 60

QY 61 SDARKSETTITQSGYNGADVGQADNSTIETQNGFRNNATYDQLVTRVTHVMAHAGG 120  
DB 61 SDARKSETTITQSGYNGADVGQADNSTIETQNGFRNNATIDQWNAKNSDITVGYGG 120  
QY 121 NNAALVNOTASDSSVMVRQVGFNNATANQY 151  
DB 121 NNAALVNDQLVTRVTHVMAHANNATANQY 151  
RESULT 10  
AAB36353  
ID AAB36353 standard; protein; 151 AA.  
XX  
AC AAB36353;  
XX  
DT 26-FEB-2001 (first entry)  
XX  
DE Agfa::PT3#8 amino acid sequence SEQ ID NO:26.  
XX  
KW Salmonella; agfa; chromosomal gene replacement; fimbrin; epitope;  
XX vaccine; immune response; immunogen.  
XX  
OS Salmonella enteritidis.  
OS Escherichia coli.  
OS Synthetic.  
XX  
FN WO200060102-A2.  
XX  
PD 12-OCT-2000.  
XX  
PF 05-APR-2000; 2000WO-CA000356.  
XX  
PR 05-APR-1999; 99US-0127888P.  
XX  
PA (UYVI-) UNIV VICTORIA.  
XX  
PI White AP, Doran JL, Collison SK, Kay WW;  
XX  
DR WPI; 2000-672631/65.  
DR N-PSDB; AAC64629.  
XX  
PT Recombinant agfa gene having a segment replaced by a foreign DNA sequence  
PT which encodes foreign epitope or antigen, expresses recombinant Agfa  
PT protein useful for eliciting immune response in animal.  
XX  
PS Disclosure; Page 138; 139pp; English.  
XX  
CC The present invention describes a recombinant agfa gene (I) where a  
CC segment of the gene has been replaced by a segment of a foreign DNA  
CC sequence which encodes a foreign epitope or antigen. Also described are:  
CC (1) use of thin aggregative fimbriae (SEF17/TAF) nucleation depended  
CC assembly system of strains of Salmonella, Escherichia coli and  
CC Enterobacteriaceae for the production of fimbriae comprising recombinant  
CC Agfa, CsgA and Agfa-homologue fimbria subunits, respectively; (2)  
CC directing recombination of a recombinant gene into the chromosome of the  
CC homologous species; (3) directing recombination of a recombinant gene  
CC back into the chromosome of the homologous species, replacing the native  
CC copy of that gene; and (4) eliciting an immune response in an animal,  
CC comprising separating an amino acid polymer comprising a recombinant Agfa  
CC protein containing a replacement segment or segments of foreign amino  
CC acid sequence or sequences grown on a Salmonella, E. coli or  
CC Enterobacteriaceae host cell, from the host cell and introducing the  
CC polymer into the animal in conjunction with a carrier or diluent. (I) is  
CC useful for the expression of recombinant Agfa protein which is useful for  
CC eliciting an immune response in an animal. In a fimbrial presentation  
CC system the heterologous antigens are presented in high numbers (up to  
CC 500,000 copies/cell), the hybrid fimbria protein possesses both the  
CC immunogenicity and adhesion properties relevant for an efficient live  
CC vaccine, the carrier fimbrial subunit proteins are usually strong  
CC immunogens, which may be important for directing an immune response  
CC against the inserted epitope, and hybrid fimbriae are easy and  
CC inexpensive to purify in large amount. The present sequence is given in  
CC the exemplification of the present invention  
XX  
SQ Sequence 151 AA;  
Query Match 78.5%; Score 603; DB 3; Length 151;  
Best Local Similarity 80.8%; Pred. No. 1.7e-50;  
Matches 122; Conservative 6; Mismatches 23; Indels 0; Gaps 0;  
1 MKLLKVAFAAIVVSGSALAGVVPQWGGGNGHNGGNSGPDSTLSIYQGSANAALALQ 60  
1 MKLLKVAFAAIVVSGSALAGVVPQWGGGNGHNGGNSGPDSTLSIYQGSANAALALQ 60



CC (1) use of thin aggregative fimbriae (SBE17/TAF) nucleation depended  
 CC assembly system of strains of Salmonella, Escherichia coli and  
 CC Enterobacteriaceae for the production of fimbriae comprising recombinant  
 CC AgfA, CsgA and AgfA-homologue fimbrial subunits, respectively; (2)  
 CC directing recombination of a recombinant gene into the chromosome of the  
 CC homologous species; (3) directing recombination of a recombinant gene  
 CC back into the chromosome of the homologous species, replacing the native  
 CC copy of that gene; and (4) eliciting an immune response in an animal,  
 CC comprising separating an amino acid polymer or segments of a recombinant AgfA  
 CC protein containing a replacement segment or segments of foreign amino  
 CC acid sequence or sequences grown on a Salmonella, E. coli or  
 CC Enterobacteriaceae host cell, from the host cell and introducing the  
 CC polymer into the animal in conjunction with a carrier or diluent. (1) is  
 CC useful for the expression of recombinant AgfA protein which is useful for  
 CC eliciting an immune response in an animal. In a fimbrial presentation  
 CC system the heterologous antigens are presented in high numbers (up to  
 CC 500,000 copies/cell), the hybrid fimbrial protein possesses both the  
 CC immunogenicity and adhesion properties relevant for an efficient live  
 CC vaccine, the carrier fimbrial subunit proteins are usually strong  
 CC immunogens, which may be important for directing an immune response  
 CC against the inserted epitope, and hybrid fimbriae are easy and  
 CC inexpensive to purify in large amount. The present sequence is given in  
 CC the exemplification of the present invention

XX Sequence 151 AA;

Query Match 77.7%; Score 597; DB 3; Length 151;  
 Best Local Similarity 81.5%; Pred. No. 6.6e-50;  
 Matches 123; Conservative 5; Mismatches 23; Indels 0; Gaps 0;  
 QY 1 MKLLKVAFAAIVVSGSALAGVVPQWGGGNGGNSGGPSTLSIYQYGSANAALALQ 60  
 Db 1 MKLLKVAFAAIVVSGSALAGVVPQWGGGNGGNSGGPSTLSIYQYGSANAALALQ 60  
 QY 61 SDARKSETTITQSGYNGADVGQADNSTIELTONGFRNNATYDQLVTRVVTHEMAHAGG 120  
 Db 61 SDARKSETTITQSGYNGADVGQADNSTIELTONGFRNNATYDQLVTRVVTHEMAHAGG 120

RESULT 13

ID AAB36348  
 XX AAB36348 standard; protein; 151 AA.  
 AC AAB36348;  
 XX 26-FEB-2001 (first entry)  
 DE AgfA::PT3#3 amino acid sequence SEQ ID NO:16.

XX Salmonella; agfA; chromosomal gene replacement; fimbrin; epitope;  
 KW vaccine; immune response; immunogen.

XX Salmonella enteritidis.

OS Escherichia coli.

OS Synthetic.

XX WO2000060102-A2.

XX 12-OCT-2000.

XX 05-APR-2000; 2000WO-CA000356.

XX 05-APR-1999; 99US-0127888P.

XX (UUVI-) UNIV VICTORIA.

XX White AP, Doran JL, Collison SK, Kay WH;

XX WPI; 2000-672631/65.

DR N-PSDB; AAC64624.  
 XX Recombinant agfA gene having a segment replaced by a foreign DNA sequence  
 PT which encodes foreign epitope or antigen, expresses recombinant AgfA  
 PT protein useful for eliciting immune response in animal.  
 XX Disclosure; Page 136; 139pp; English.

XX The present invention describes a recombinant agfA gene (1) where a  
 CC segment of the gene has been replaced by a segment of a foreign DNA  
 CC sequence which encodes a foreign epitope or antigen. Also described are:  
 CC (1) use of thin aggregative fimbriae (SBE17/TAF) nucleation depended  
 CC assembly system of strains of Salmonella, Escherichia coli and  
 CC Enterobacteriaceae for the production of fimbriae comprising recombinant  
 CC AgfA, CsgA and AgfA-homologue fimbrial subunits, respectively; (2)  
 CC directing recombination of a recombinant gene into the chromosome of the  
 CC homologous species; (3) directing recombination of a recombinant gene  
 CC back into the chromosome of the homologous species, replacing the native  
 CC copy of that gene; and (4) eliciting an immune response in an animal,  
 CC comprising separating an amino acid polymer or segments of foreign amino  
 CC protein containing a replacement segment or segments of foreign amino  
 CC acid sequence or sequences grown on a Salmonella, E. coli or  
 CC Enterobacteriaceae host cell, from the host cell and introducing the  
 CC polymer into the animal in conjunction with a carrier or diluent. (1) is  
 CC useful for the expression of recombinant AgfA protein which is useful for  
 CC eliciting an immune response in an animal. In a fimbrial presentation  
 CC system the heterologous antigens are presented in high numbers (up to  
 CC 500,000 copies/cell), the hybrid fimbrial protein possesses both the  
 CC immunogenicity and adhesion properties relevant for an efficient live  
 CC vaccine, the carrier fimbrial subunit proteins are usually strong  
 CC immunogens, which may be important for directing an immune response  
 CC against the inserted epitope, and hybrid fimbriae are easy and  
 CC inexpensive to purify in large amount. The present sequence is given in  
 CC the exemplification of the present invention

XX Sequence 151 AA;

Query Match 73.7%; Score 566; DB 3; Length 151;  
 Best Local Similarity 80.8%; Pred. No. 6.7e-47;  
 Matches 122; Conservative 5; Mismatches 24; Indels 0; Gaps 0;

QY 1 MKLLKVAFAAIVVSGSALAGVVPQWGGGNGGNSGGPSTLSIYQYGSANAALALQ 60  
 Db 1 MKLLKVAFAAIVVSGSALAGVVPQWGGGNGGNSGGPSTLSIYQYGSANAALALQ 60  
 QY 61 SDARKSETTITQSGYNGADVGQADNSTIELTONGFRNNATYDQLVTRVVTHEMAHAGG 120  
 Db 61 SDARKSETTITQSGYNGADVGQADNSTIELTONGFRNNATYDQLVTRVVTHEMAHAGG 120  
 QY 121 NNAALVNQATSDSSVMVRQVGFNNATANQY 151  
 Db 121 NNAALVNQATSDSSVMVRQVGFNNATANQY 151

RESULT 14

ID AAB36343  
 XX AAB36343 standard; protein; 151 AA.

AC AAB36343;

XX 26-FEB-2001 (first entry)

XX Escherichia coli CsgA amino acid sequence SEQ ID NO:7.

XX Salmonella; agfA; chromosomal gene replacement; fimbrin; epitope;  
 KW vaccine; immune response; immunogen.

OS Escherichia coli.

XX WO2000060102-A2.

XX 12-OCT-2000.

05-APR-2000; 2000WO-CA000356.  
 05-APR-1999; 99US-0127888P.  
 (UYVI-) UNIV VICTORIA.  
 White AP, Doran JL, Collison SK, Kay WW;  
 WPI; 2000-672631/65.  
 N-PSDB; AAC64619.  
 Recombinant agfa gene having a segment replaced by a foreign DNA sequence  
 PT which encodes foreign epitope or antigen, expresses recombinant Agfa  
 protein useful for eliciting immune response in animal.  
 XX  
 PS Disclosure; Page 135; 139pp; English.  
 XX  
 CC The present invention describes a recombinant agfa gene (I) where a  
 CC segment of the gene has been replaced by a segment of a foreign DNA  
 CC sequence which encodes a foreign epitope or antigen. Also described are:  
 CC (1) use of thin aggregative fimbriae (SEPI7/7AF) nucleation depended  
 CC assembly system of strains of Salmonella, Escherichia coli and  
 CC Enterobacteriaceae for the production of fimbriae comprising recombinant  
 CC Agfa, CsgA and Agfa-homologue fimbrial subunits, respectively; (2)  
 CC directing recombination of a recombinant gene into the chromosome of the  
 CC homologous species; (3) directing recombination of a recombinant gene  
 CC back into the chromosome of the homologous species, replacing the native  
 CC copy of that gene; and (4) eliciting an immune response in an animal,  
 CC comprising separating an amino acid polymer comprising a recombinant Agfa  
 CC protein containing a replacement segment or segments of foreign amino  
 CC acid sequence or sequences grown on a Salmonella, E. coli or  
 CC Enterobacteriaceae host cell, from the host cell and introducing the  
 CC polymer into the animal in conjunction with a carrier or diluent. (I) is  
 CC useful for the expression of recombinant Agfa protein which is useful for  
 CC eliciting an immune response in an animal. In a fimbrial presentation  
 CC system the heterologous antigens are presented in high numbers (up to  
 CC 500,000 copies/cell), the hybrid fimbrial protein possesses both the  
 CC immunogenicity and adhesion properties relevant for an efficient live  
 CC vaccine, the carrier fimbrial subunit proteins are usually strong  
 CC immunogens, which may be important for directing an immune response  
 CC against the inserted epitope, and hybrid fimbriae are easy and  
 CC inexpensive to purify in large amount. The present sequence is given in  
 CC the exemplification of the present invention  
 XX  
 SQ Sequence 151 AA;  
 Query Match 67.3%; Score 517; DB 3; Length 151;  
 Best Local Similarity 69.5%; Pred. No. 3.8e-42;  
 Matches 105; Conservative 17; Mismatches 29; Indels 0; Gaps 0;  
 QY 1 MKLLKVAFAAIVVSGSALAGVVPQGGGNGHNGGNSGPDSTLSIYQGSANAALALQ 60  
 DB 1 MKLLKVAFAAIVVSGSALAGVVPQGGGNGHNGGNSGPDSTLSIYQGSANAALALQ 60  
 QY 61 SDARKSETTITQSGYNGADVGGADNNTIELTQNGFRNNATYDQLVTRVVTHEMAHAGG 120  
 DB 61 TDARNSDLTITQHGGNGADVGGSDSSIDLITQRFNGSATLDQWNGKNSMTVKQFGG 120  
 QY 121 NNAALVNQTASDSSVMVRQVGFNNATANQY 151  
 DB 121 GNGAAVDQTASNSNVNTQVGFNNATAHQY 151  
 Search completed: August 2, 2004, 14:48:29  
 Job time : 44.9 secs

05-APR-2000; 2000WO-CA000356.  
 05-APR-1999; 99US-0127888P.  
 (UYVI-) UNIV VICTORIA.  
 White AP, Doran JL, Collison SK, Kay WW;  
 WPI; 2000-672631/65.  
 N-PSDB; AAC64619.  
 Recombinant agfa gene having a segment replaced by a foreign DNA sequence  
 PT which encodes foreign epitope or antigen, expresses recombinant Agfa  
 protein useful for eliciting immune response in animal.  
 XX  
 PS Disclosure; Page 135; 139pp; English.  
 XX  
 CC The present invention describes a recombinant agfa gene (I) where a  
 CC segment of the gene has been replaced by a segment of a foreign DNA  
 CC sequence which encodes a foreign epitope or antigen. Also described are:  
 CC (1) use of thin aggregative fimbriae (SEPI7/7AF) nucleation depended  
 CC assembly system of strains of Salmonella, Escherichia coli and  
 CC Enterobacteriaceae for the production of fimbriae comprising recombinant  
 CC Agfa, CsgA and Agfa-homologue fimbrial subunits, respectively; (2)  
 CC directing recombination of a recombinant gene into the chromosome of the  
 CC homologous species; (3) directing recombination of a recombinant gene  
 CC back into the chromosome of the homologous species, replacing the native  
 CC copy of that gene; and (4) eliciting an immune response in an animal,  
 CC comprising separating an amino acid polymer comprising a recombinant Agfa  
 CC protein containing a replacement segment or segments of foreign amino  
 CC acid sequence or sequences grown on a Salmonella, E. coli or  
 CC Enterobacteriaceae host cell, from the host cell and introducing the  
 CC polymer into the animal in conjunction with a carrier or diluent. (I) is  
 CC useful for the expression of recombinant Agfa protein which is useful for  
 CC eliciting an immune response in an animal. In a fimbrial presentation  
 CC system the heterologous antigens are presented in high numbers (up to  
 CC 500,000 copies/cell), the hybrid fimbrial protein possesses both the  
 CC immunogenicity and adhesion properties relevant for an efficient live  
 CC vaccine, the carrier fimbrial subunit proteins are usually strong  
 CC immunogens, which may be important for directing an immune response  
 CC against the inserted epitope, and hybrid fimbriae are easy and  
 CC inexpensive to purify in large amount. The present sequence is given in  
 CC the exemplification of the present invention  
 XX  
 SQ Sequence 151 AA;  
 Query Match 67.3%; Score 517; DB 3; Length 151;  
 Best Local Similarity 69.5%; Pred. No. 3.8e-42;  
 Matches 105; Conservative 17; Mismatches 29; Indels 0; Gaps 0;  
 QY 1 MKLLKVAFAAIVVSGSALAGVVPQGGGNGHNGGNSGPDSTLSIYQGSANAALALQ 60  
 DB 1 MKLLKVAFAAIVVSGSALAGVVPQGGGNGHNGGNSGPDSTLSIYQGSANAALALQ 60  
 QY 61 SDARKSETTITQSGYNGADVGGADNNTIELTQNGFRNNATYDQLVTRVVTHEMAHAGG 120  
 DB 61 TDARNSDLTITQHGGNGADVGGSDSSIDLITQRFNGSATLDQWNGKNSMTVKQFGG 120  
 QY 121 NNAALVNQTASDSSVMVRQVGFNNATANQY 151  
 DB 121 GNGAAVDQTASNSNVNTQVGFNNATAHQY 151  
 RESULT 15  
 ABR82651  
 ID ABR82651 standard; protein; 151 AA.  
 XX  
 AC ABR82651;  
 XX  
 DT 04-DEC-2003 (first entry)  
 XX  
 DE E. coli CsgA subunit 15 kDa protein.  
 XX

Plasma protein; immune response; antibacterial; vaccine; gene therapy.  
 XX  
 OS Escherichia coli.  
 XX  
 PN WO2003064446-A2.  
 XX  
 PD 07-AUG-2003.  
 XX  
 PF 30-JAN-2003; 2003WO-EP000943.  
 XX  
 PR 31-JAN-2002; 2002GB-00002275.  
 XX  
 PA (HANS-) HANSA MEDICAL RES AB.  
 XX  
 PI Bjoerck L, Olsen A, Wikstroem M, Herwald H;  
 XX  
 DR WPI; 2003-646136/61.  
 DR N-PSDB; ACF36153.  
 XX  
 PT New isolated peptide capable of binding a mammalian plasma protein,  
 PT useful in the manufacture of a medicament for the prevention and/or  
 PT treatment of a bacterial infection, such as Escherichia coli, Salmonella  
 PT or Shigella infections.  
 XX  
 PS Disclosure; Page 41-42; 42pp; English.  
 XX  
 CC The invention relates to an isolated peptide capable of binding a  
 CC mammalian plasma protein or of generating an immune response in a mammal  
 CC selected from sequences shown in ABR82642, ABR82648-49. The peptide or  
 CC antibody is useful for treating a bacterial infection in a human or  
 CC animal or in the manufacture of a medicament for the prophylactic  
 CC treatment of a bacterial infection, such as Escherichia coli, Salmonella  
 CC or Shigella infection. The peptide that is immobilized on a solid support  
 CC is also useful as a reagent for determining the ability of a plasma  
 CC protein to bind to bacteria. The present sequence represents an E. coli  
 CC 15 kDa protein  
 XX  
 SQ Sequence 151 AA;  
 Query Match 66.7%; Score 512; DB 7; Length 151;  
 Best Local Similarity 68.9%; Pred. No. 1.2e-41;  
 Matches 104; Conservative 17; Mismatches 30; Indels 0; Gaps 0;  
 QY 1 MKLLKVAFAAIVVSGSALAGVVPQGGGNGHNGGNSGPDSTLSIYQGSANAALALQ 60  
 DB 1 MKLLKVAFAAIVVSGSALAGVVPQGGGNGHNGGNSGPDSTLSIYQGSANAALALQ 60  
 QY 61 SDARKSETTITQSGYNGADVGGADNNTIELTQNGFRNNATYDQLVTRVVTHEMAHAGG 120  
 DB 61 TDARNSDLTITQHGGNGADVGGSDSSIDLITQRFNGSATLDQWNGKNSMTVKQFGG 120  
 QY 121 NNAALVNQTASDSSVMVRQVGFNNATANQY 151  
 DB 121 GNGAAVDQTASNSNVNTQVGFNNATAHQY 151  
 Search completed: August 2, 2004, 14:48:29  
 Job time : 44.9 secs



GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: August 2, 2004, 14:40:48 ; Search time 12 Seconds  
(without alignments)  
649.627 Million cell updates/sec

Title: US-09-543-407-30  
Perfect score: 768  
Sequence: 1 MKLLKVAFAAIVVSGSALA.....DSVMVROVGFNNATANQY 151

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 389414 seqs, 51625971 residues

Total number of hits satisfying chosen parameters: 389414

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Issued Patents AA:\*  
1: /cgn2\_6/ptodata/2/iaa/5A\_COMB.pep:\*  
2: /cgn2\_6/ptodata/2/iaa/5B\_COMB.pep:\*  
3: /cgn2\_6/ptodata/2/iaa/6A\_COMB.pep:\*  
4: /cgn2\_6/ptodata/2/iaa/6B\_COMB.pep:\*  
5: /cgn2\_6/ptodata/2/iaa/PCTUS\_COMB.pep:\*  
6: /cgn2\_6/ptodata/2/iaa/backfiles1.pep:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	676	88.0	151	1	US-08-233-788A-59 Sequence 59, Appl
2	496	64.6	120	1	US-08-233-788A-57 Sequence 57, Appl
3	92.5	12.0	738	3	US-08-864-038A-3 Sequence 3, Appl
4	91.5	11.9	673	3	US-09-196-387-8 Sequence 8, Appl
5	91.5	11.9	673	4	US-09-841-835-8 Sequence 8, Appl
6	91.5	11.9	949	3	US-09-196-387-10 Sequence 10, Appl
7	91.5	11.9	949	4	US-09-841-835-10 Sequence 10, Appl
8	91.5	11.9	1327	3	US-09-196-387-2 Sequence 2, Appl
9	91.5	11.9	1327	4	US-09-841-835-2 Sequence 2, Appl
10	91.5	11.9	1327	4	US-09-972-115A-8 Sequence 8, Appl
11	86.5	11.3	745	4	US-09-336-115C-6 Sequence 6, Appl
12	84	10.9	2736	4	US-09-336-115C-6 Sequence 6, Appl
13	83	10.8	1216	4	US-09-252-991A-30227 Sequence 30227, A
14	81.5	10.6	589	4	US-09-134-000C-5130 Sequence 5130, Ap
15	80	10.4	867	4	US-09-489-039A-7849 Sequence 7849, Ap
16	80	10.4	878	4	US-09-540-236-2676 Sequence 2676, Ap
17	78.5	10.2	309	4	US-09-540-236-3401 Sequence 3401, Ap
18	78.5	10.2	873	4	US-09-252-991A-22266 Sequence 22266, A
19	78.5	10.2	892	4	US-09-336-447A-13 Sequence 13, Appl
20	78.5	10.2	1415	4	US-09-336-447A-5 Sequence 5, Appl
21	77.5	10.1	1160	3	US-08-808-593A-24 Sequence 24, Appl
22	77.5	10.1	1207	4	US-09-252-991A-26438 Sequence 26438, A
23	77	10.0	415	4	US-09-025-769B-11518 Sequence 11518, A
24	77	10.0	975	4	US-09-328-352-4764 Sequence 4764, Ap
25	77	10.0	1003	1	US-08-571-758-4 Sequence 4, Appl
26	77	10.0	1003	1	US-08-909-984A-4 Sequence 4, Appl
27	77	10.0	1003	1	US-08-909-983-4 Sequence 4, Appl

28	77	10.0	1034	4	US-09-252-991A-26658 Sequence 26658, A
29	77	10.0	1690	4	US-09-595-684B-39 Sequence 39, Appl
30	77	10.0	1864	2	US-08-804-227C-3 Sequence 3, Appl
31	77	10.0	2315	4	US-09-543-681A-5434 Sequence 5434, Ap
32	76.5	10.0	624	4	US-09-336-447A-9 Sequence 9, Appl
33	76.5	10.0	941	4	US-09-336-447A-9 Sequence 9, Appl
34	76.5	10.0	1398	1	US-08-750-532-9 Sequence 9, Appl
35	76.5	10.0	1398	3	US-08-894-818B-8 Sequence 8, Appl
36	76.5	10.0	1398	4	US-09-445-472-6 Sequence 6, Appl
37	76.5	10.0	1612	1	US-08-169-927-2 Sequence 2, Appl
38	76.5	10.0	2123	3	US-08-968-685A-10 Sequence 10, Appl
39	76	9.9	231	4	US-09-540-236-3827 Sequence 3827, Ap
40	76	9.9	870	4	US-09-177-850-91 Sequence 91, Appl
41	75.5	9.8	238	4	US-09-495-880A-42 Sequence 42, Appl
42	75.5	9.8	282	2	US-08-860-174A-10 Sequence 10, Appl
43	75.5	9.8	365	4	US-09-489-847-322 Sequence 322, App
44	75.5	9.8	394	4	US-09-252-991A-28148 Sequence 28148, A
45	75.5	9.8	432	4	US-09-403-089A-1 Sequence 1, Appl

ALIGNMENTS

RESULT 1  
US-08-233-788A-59  
; Sequence 59, Application US/08233788A  
; Patent No. 5635617  
; GENERAL INFORMATION:  
; APPLICANT: Doran, James L.  
; APPLICANT: Kay, William W.  
; APPLICANT: Collinson, Karen S.  
; APPLICANT: Clouthier, Sharon C.  
; TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR DETECTION  
; TITLE OF INVENTION: OF SALMONELLA  
; NUMBER OF SEQUENCES: 61  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Seed and Berry  
; STREET: 6300 Columbia Center, 701 Fifth Avenue  
; CITY: Seattle  
; STATE: Washington  
; COUNTRY: U.S.A.  
; ZIP: 98104-7092  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patentin Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/233,788A  
; FILING DATE: 26-APR-1994  
; CLASSIFICATION: 435  
; ATTORNEY/AGENT INFORMATION:  
; NAME: King, Joshua  
; REGISTRATION NUMBER: 35,570  
; REFERENCE/DOCKET NUMBER: 920043.403C2  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (206) 682-4900  
; TELEFAX: (206) 682-6031  
; TELEX: 3723836 SEEDANBERRY  
; INFORMATION FOR SEQ ID NO: 59:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 151 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
US-08-233-788A-59

Query Match 88.0%; Score 676; DB 1; Length 151;  
Best Local Similarity 90.1%; Pred. No. 2.7e-61;  
Matches 136; Conservative 2; Mismatches 13; Indels 0;  
Gaps 0;  
1 MKLLKVAFAAIVVSGSALGVFOWGGGNNHNGSSGPDSTLSIYQVGSANAALQ 60  
|||||

Db 1 MLLKVAFAAIVVSGSALAGVVPWGCGGNGHNGSGDPDSTLSIYQGSANAALALQ 60  
QY 61 SDARKSETTITQSGYNGADVGQGGADNSTIELTQGFNNATYDQLVTRVVTTHMAHAGG 120  
Db 61 SDARKSETTITQSGYNGADVGQGGADNSTIELTQGFNNATIDWNKNSDITVGYGG 120  
QY 121 NNAALVNTASDSSVMVQVGFNNATANY 151  
Db 121 NNPALVNTASDSSVMVQVGFNNATANY 151

## RESULT 2

US-08-233-788A-57  
; Sequence 57, Application US/08233788A  
; Patent No. 5635617  
; GENERAL INFORMATION:  
; APPLICANT: Doran, James L.  
; APPLICANT: Kay, William W.  
; APPLICANT: Collinson, Karen C.  
; APPLICANT: Clouthier, Sharon C.  
; TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR DETECTION  
; TITLE OF INVENTION: OF SALMONELLA  
; NUMBER OF SEQUENCES: 61  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Seed and Berry  
; STREET: 6300 Columbia Center, 701 Fifth Avenue  
; CITY: Seattle  
; STATE: Washington  
; COUNTRY: U.S.A.  
; ZIP: 98104-7092  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patent in Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/233,788A  
; FILING DATE: 26-APR-1994  
; CLASSIFICATION: 435  
; ATTORNEY/AGENT INFORMATION:  
; NAME: King, Joshua  
; REGISTRATION NUMBER: 35,570  
; REFERENCE/DOCKET NUMBER: 920043.403C2  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (206) 622-4900  
; TELEFAX: (206) 682-6031  
; TELEX: 3723836 SEDANBERRY  
; INFORMATION FOR SEQ ID NO: 57:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 120 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
US-08-233-788A-57

Query Match 64.6%; Score 496; DB 1; Length 120;  
Best Local Similarity 87.5%; Pred. No. 4e-43;  
Matches 98; Conservative 12; Mismatches 0; Gaps 0;

QY 22 VVPWGCGGNGHNGSGDPDSTLSIYQGSANAALALQSDKSETTITQSGYNGADV 81  
Db 1 VVPWGCGGNGHNGSGDPDSTLSIYQGSANAALALQSDKSETTITQSGYNGADV 60  
QY 82 GQGDNSTIELTQGFNNATYDQLVTRVVTTHMAHAGNNAALVNTASDS 133  
Db 61 GQGDNSTIELTQGFNNATIDWNKNSDITVGYGGNNAALVNTASDS 112

## RESULT 3

US-08-864-038A-3  
; Sequence 3, Application US/08864038A  
; Patent No. 6001592  
; GENERAL INFORMATION:

; APPLICANT: Kunio NAKASHIMA et al.  
; TITLE OF INVENTION: NOVEL POLYPEPTIDE GENE CDNA, VECTOR  
; TITLE OF INVENTION: CONTAINING SAID CDNA, HOST CELLS TRANSFORMED WITH SAID  
; TITLE OF INVENTION: VECTOR, POLYPEPTIDE PRODUCED THEREBY, METHOD OF PRODUCING  
; TITLE OF INVENTION: SAID POLYPEPTIDE, DNA ENCODING SAID POLYPEPTIDE AND ANTIBODY  
; NUMBER OF SEQUENCES: 4  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: 812-5 Hirano  
; STREET: Isshinden  
; CITY: Tsu-city  
; STATE: Mie-prefecture  
; COUNTRY: JAPAN  
; ZIP: 514-01  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Diskette, 3.50 inch, 1.44 MB storage  
; COMPUTER: IBM Compatible  
; OPERATING SYSTEM: Microsoft Windows 95  
; SOFTWARE: Word Perfect 6.1  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/864,038A  
; FILING DATE: May 28, 1997  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: JP 8-184459  
; FILING DATE: 15-July-1996  
; ATTORNEY/AGENT INFORMATION:  
; NAME: C. Bruce Hamburg  
; REGISTRATION NUMBER: 22,389  
; REFERENCE/DOCKET NUMBER: F-5610  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (212) 986-2340  
; TELEFAX: (212) 953-7733  
; INFORMATION FOR SEQ ID NO: 3:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 738  
; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
; ORIGINAL SOURCE:  
; ORGANISM: Pinctada fucata  
; CELL TYPE: mantle epithelial cell  
; FEATURE:  
; NAME/KEY: peptide  
; LOCATION: from 1 to 738  
; IDENTIFICATION METHOD: E (by experiment)  
US-08-864-038A-3

Query Match 12.0%; Score 92.5; DB 3; Length 738;  
Best Local Similarity 27.7%; Pred. No. 0.46;  
Matches 43; Conservative 12; Mismatches 61; Indels 39; Gaps 6;

QY 3 LLKVAFAAIVVSGSALAGVVPWGCGGNGHNGSGDPDSTLSIYQGSANAALALQSD 62  
Db 419 LLKSSASASASASASAG-----GGGGGNGGNGGGG-----GGGAGALA----- 460  
QY 63 ARKSETTITQSGYNGADVGQGGADNSTIELTQGFNNATYDQLVTRVVTTHMAHAGN 122  
Db 461 -----AALAAAGAGGLGGGGGAGALAAALAAAG-AGGGGFGGL-----GGGLGLGGGS 508  
QY 123 AALVNTASDSS-----VMVQVGFNNATA 148  
Db 509 AAAAAAASGGGGRALRRRLRRMRGGGSAAA 543

## RESULT 4

US-09-196-387-8  
; Sequence 8, Application US/09196387  
; Patent No. 6277613  
; GENERAL INFORMATION:  
; APPLICANT: de Lange, Titia  
; APPLICANT: Smith, Susan  
; TITLE OF INVENTION: A PROTEIN THAT BINDS TO TRF1 AND METHODS  
; OF USE THEREOF



```

; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/841.835
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 09/196,387
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Jackson Esq., David A.
; REGISTRATION NUMBER: 26,742
; REFERENCE/DOCKET NUMBER: 600-1-230 CIP1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 201-487-5800
; TELEFAX: 201-343-1684
; TELEX: 133521
; INFORMATION FOR SEQ ID NO: 8:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 673 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-09-841-835-8

Query Match 11.9%; Score 91.5; DB 4; Length 673;
Best Local Similarity 29.4%; Pred.No.0.51;
Matches 42; Conservative 17; Mismatches 65; Indels 19; Gaps 6

QY 6 VAAFAAI-VYSGALAGVWPQWGGNHNHGGNSGPDSTLSIYQYGSANAALALQSDAR 64
DB 99 VAAAPVPAVSTSSAACGAPNPAGSGNNGNPPSSSSPTSS-SSSSPSPGSSLAESPEAA 157
QY 65 KSETTIT----QSGYNGADVGQADNSTLTLTQNGFRNNTATYDQLVTRVVTHEMAHAGG 120
DB 158 GVSTAPLPGGAAGPGTGVPAVSGALKEELLEACRNGD-----VSRV--KRLVDAAN 206
QY 121 NNAALVNQTASDSSVMVRQVGF 143
DB 207 VNAK--DMAGKSSPLHFAAGFG 227

RESULT 6
US-09-196-387-10
; Sequence 10, Application US/09196387
; Patent No. 6277613
; GENERAL INFORMATION:
; APPLICANT: de Lange, Titia
; APPLICANT: Smith, Susan
; TITLE OF INVENTION: A PROTEIN THAT BINDS TO TRF1 AND METHODS
; TITLE OF INVENTION: OF USE THEREOF
; NUMBER OF SEQUENCES: 12
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Klauber & Jackson
; STREET: 411 Hackensack Avenue, 4th Floor
; CITY: Hackensack
; STATE: New Jersey
; COUNTRY: USA
; ZIP: 07601
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/196,387
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 09/095,225
; FILING DATE: June 10, 1998
; ATTORNEY/AGENT INFORMATION:
; NAME: Jackson Esq., David A.

```

```
;
; REGISTRATION NUMBER: 26,742
; REFERENCE/DOCKET NUMBER: 600-1-230 CIP1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 201-487-5800
; TELEFAX: 201-343-1684
; TELEX: 133521
; INFORMATION FOR SEQ ID NO: 10:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 949 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-09-196-387-10

Query Match 11.9%; Score 91.5; DB 3; Length 949;
Best Local Similarity 29.4%; Pred. No. 0.8;
Matches 42; Conservative 17; Mismatches 65; Indels 19; Gaps 6;

QY 6 VAAFAAI-VVSGSALAGVVPQWGGGNGGNGSGGPDSTLSIYQYGSANAALALQSDAR 64
DB 99 VAAAPVVPVAVSTSSAAGVAPNPGSGNNSSSSPTSS-SSSPSPSGSLAESPEAA 157
QY 65 KSETTIT-----OSGYNGADVCGQADNSTIETQNGFRNNATYDOLVTRVVTHEMAHAGG 120
DB 158 GVSSTAPLPGGAGPGTGPVAVSGALRELEACRNGD-----VSRV--KRLVDAAN 206
QY 121 NNAALVNQATSDSSVMVRQVFG 143
DB 207 VNAK--DMAGRKSSPLHFAAGFG 227

RESULT 7
US-09-841-835-10
; Sequence 10, Application US/09841835
; Patent No. 6506587
; GENERAL INFORMATION:
; APPLICANT: de Lange, Titia
; APPLICANT: Smith, Susan
; TITLE OF INVENTION: A PROTEIN THAT BINDS TO TRP1 AND METHODS
; TITLE OF INVENTION: OF USE THEREOF
; NUMBER OF SEQUENCES: 12
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Klauber & Jackson
; STREET: 411 Hackensack Avenue, 4th Floor
; CITY: Hackensack
; STATE: New Jersey
; COUNTRY: USA
; ZIP: 07601
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent in Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/841,835
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 09/095,225
; FILING DATE: June 10, 1998
; ATTORNEY/AGENT INFORMATION:
; NAME: Jackson Esq., David A.
; REGISTRATION NUMBER: 26,742
; REFERENCE/DOCKET NUMBER: 600-1-230 CIP1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 201-487-5800
; TELEFAX: 201-343-1684
; TELEX: 133521
; INFORMATION FOR SEQ ID NO: 10:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 949 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; HYPOTHETICAL: NO
; US-09-196-387-2

Query Match 11.9%; Score 91.5; DB 3; Length 1327;
Best Local Similarity 29.4%; Pred. No. 1.3;
Matches 42; Conservative 17; Mismatches 65; Indels 19; Gaps 6;

QY 6 VAAFAAI-VVSGSALAGVVPQWGGGNGGNGSGGPDSTLSIYQYGSANAALALQSDAR 64
DB 99 VAAAPVVPVAVSTSSAAGVAPNPGSGNNSSSSPTSS-SSSPSPSGSLAESPEAA 157
QY 65 KSETTIT-----OSGYNGADVCGQADNSTIETQNGFRNNATYDOLVTRVVTHEMAHAGG 120
DB 158 GVSSTAPLPGGAGPGTGPVAVSGALRELEACRNGD-----VSRV--KRLVDAAN 206
QY 121 NNAALVNQATSDSSVMVRQVFG 143
DB 207 VNAK--DMAGRKSSPLHFAAGFG 227

RESULT 8
US-09-196-387-2
; Sequence 2, Application US/09196387
; Patent No. 6277613
; GENERAL INFORMATION:
; APPLICANT: de Lange, Titia
; APPLICANT: Smith, Susan
; TITLE OF INVENTION: A PROTEIN THAT BINDS TO TRP1 AND METHODS
; TITLE OF INVENTION: OF USE THEREOF
; NUMBER OF SEQUENCES: 12
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Klauber & Jackson
; STREET: 411 Hackensack Avenue, 4th Floor
; CITY: Hackensack
; STATE: New Jersey
; COUNTRY: USA
; ZIP: 07601
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent in Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/196,387
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 09/095,225
; FILING DATE: June 10, 1998
; ATTORNEY/AGENT INFORMATION:
; NAME: Jackson Esq., David A.
; REGISTRATION NUMBER: 26,742
; REFERENCE/DOCKET NUMBER: 600-1-230 CIP1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 201-487-5800
; TELEFAX: 201-343-1684
; TELEX: 133521
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1327 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; HYPOTHETICAL: NO
; US-09-196-387-2

Query Match 11.9%; Score 91.5; DB 3; Length 1327;
Best Local Similarity 29.4%; Pred. No. 1.3;
Matches 42; Conservative 17; Mismatches 65; Indels 19; Gaps 6;

QY 6 VAAFAAI-VVSGSALAGVVPQWGGGNGGNGSGGPDSTLSIYQYGSANAALALQSDAR 64
DB 99 VAAAPVVPVAVSTSSAAGVAPNPGSGNNSSSSPTSS-SSSPSPSGSLAESPEAA 157
QY 65 KSETTIT-----OSGYNGADVCGQADNSTIETQNGFRNNATYDOLVTRVVTHEMAHAGG 120
DB 158 GVSSTAPLPGGAGPGTGPVAVSGALRELEACRNGD-----VSRV--KRLVDAAN 206
QY 121 NNAALVNQATSDSSVMVRQVFG 143
DB 207 VNAK--DMAGRKSSPLHFAAGFG 227
```

```

Db      99 VAAAPVVPVSTSSAACVAPNPAGSGSNNPSSSSPTSS--SSSSPSPSGSLAESPEAA 15
QY      65 KSETTIT---OSGYNGADVCGQADNSTIELTQNGFRNNATYDQLVTRVVTTHMAHAG 120
Db      158 GVSSTAPLPGGAAGPCTGVPVAVSGALRELLACRNGD-----VSRV--KRLVDAAN 206
QY      121 NNAALVNQTSADSSVMVRQVGF 143
Db      207 VNAK--DMAGRKSSPLHFAAGF 227

RESULT 9
US-09-841-835-2
; Sequence 2, Application US/09841835
; Patent No. 6506587
; GENERAL INFORMATION:
; APPLICANT: de Lange, Titia
; APPLICANT: Smith, Susan
; TITLE OF INVENTION: A PROTEIN THAT BINDS TO TRF1 AND METHODS
; TITLE OF INVENTION: OF USE THEREOF
; NUMBER OF SEQUENCES: 12
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Klauber & Jackson
; STREET: 411 Hackensack Avenue, 4th Floor
; CITY: Hackensack
; STATE: New Jersey
; COUNTRY: USA
; ZIP: 07601
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/841,835
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 09/196,387
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Jackson Esq., David A.
; REGISTRATION NUMBER: 26,742
; REFERENCE/DOCKET NUMBER: 600-1-230 CIP1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 201-487-5800
; TELEFAX: 201-343-1684
; TELEX: 133521
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1327 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; HYPOTHETICAL: NO
; US-09-841-835-2

Query Match 11.9%; Score 91.5; DB 4; Length 1327;
Best Local Similarity 29.4%; Pred. No. 1.3;
Matches 42; Conservative 17; Mismatches 65; Indels 19; Gaps 6
QY      6 VAAFAAI-VVSGSALAGVVPQWGGNGNNGSGGPDSTLSIYQYSANAALALQSDAR 64
Db      99 VAAAPVVPVSTSSAACVAPNPAGSGSNNPSSSSPTSS--SSSSPSPSGSLAESPEAA 157
QY      65 KSETTIT---OSGYNGADVCGQADNSTIELTQNGFRNNATYDQLVTRVVTTHMAHAG 120
Db      158 GVSSTAPLPGGAAGPCTGVPVAVSGALRELLACRNGD-----VSRV--KRLVDAAN 206
QY      121 NNAALVNQTSADSSVMVRQVGF 143
Db      207 VNAK--DMAGRKSSPLHFAAGF 227

```

Query Match 11.3%; Score 86.5; DB 4; Length 745;  
 Best Local Similarity 20.0%; Pred. No. 1.9;  
 Matches 34; Conservative 27; Mismatches 72; Indels 37; Gaps 5;  
 QY 8 AFAAIVVSSALAGVVPQW-----GGGNGHNGGN----- 37  
 DB 102 AYAQVFLAINAAVGL---WNTIGYAVCMGNGNGTESGFSVIFNDPQDQSTQTCNRFE 158  
 QY 38 SSGPSTLSIYQYGSANAALQSDARKSETTITQSGYNGADYVGGADNSTIELTQNGP 97  
 DB 159 STPGKMSIDEFKLLNEAYQIIQALKNQSGFPBLG-GNGTKV---SVNYYNCEQOTAD 214  
 QY 98 RNATYDQLVTRVTHMAHAGNNAALVNQATSDSSVWVRQVFGNNAT 147  
 DB 215 INGVYQFCAKXNGSSSSNGSGSTQTATTQGGVTITTYNNKAT 264

RESULT 12  
 US-09-252-991A-30227  
 ; Sequence 30227, Application US/09252991A  
 ; Patent No. 6551795  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Marc J. Rubenfield et al.  
 ; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS  
 ; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS  
 ; FILE REFERENCE: 107196.136  
 ; CURRENT APPLICATION NUMBER: US/09/252,991A  
 ; CURRENT FILING DATE: 1999-02-18  
 ; PRIOR APPLICATION NUMBER: US 60/074,788  
 ; PRIOR FILING DATE: 1998-02-18  
 ; PRIOR APPLICATION NUMBER: US 60/094,190  
 ; PRIOR FILING DATE: 1998-07-27  
 ; NUMBER OF SEQ ID NOS: 33142  
 ; SEQ ID NO 30227  
 ; LENGTH: 2736  
 ; TYPE: PRT  
 ; ORGANISM: Pseudomonas aeruginosa  
 US-09-252-991A-30227

Query Match 10.9%; Score 84; DB 4; Length 2736;  
 Best Local Similarity 29.7%; Pred. No. 19;  
 Matches 43; Conservative 17; Mismatches 49; Indels 36; Gaps 8;  
 QY 6 VAFAAIVVSSAL-AGVV--PQWGGGNGHNGGSSGPDST-----LSIYQYGSANA 55  
 DB 149 MSAGTALLVGAAGVAGVAINSSGGG--GGGSSVPPDTPPKAASGLKIAPDGSIS 205  
 QY 56 ALALQSDARKSETTITQSGYNGADYV-GQADNSTIELTQNG-----FNNATYDQLVTR 109  
 DB 206 GQA-----EAGASVGDITNGDGKPLTVIADANGNFAPLNPPLTNGQTTV 252  
 QY 110 VVTHMAHAGNNAALVNQATSDSS 134  
 DB 253 VVT-----DPAGNASPPAQVAPDTT 273

RESULT 13  
 US-09-134-000C-5130  
 ; Sequence 5130, Application US/09134000C  
 ; Patent No. 6617156  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Lynn Doucette-Stamm et al  
 ; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO  
 ; TITLE OF INVENTION: ENTEROCOCCUS FAECALIS FOR DIAGNOSTICS AND THERAPEUTICS  
 ; FILE REFERENCE: 032796-032  
 ; CURRENT APPLICATION NUMBER: US/09/134,000C  
 ; CURRENT FILING DATE: 1998-08-13  
 ; PRIOR APPLICATION NUMBER: US 60/055,778  
 ; PRIOR FILING DATE: 1997-08-15  
 ; NUMBER OF SEQ ID NOS: 6812  
 ; SOFTWARE: Patent in version 3.1  
 ; SEQ ID NO 5130  
 ; LENGTH: 1216

TYPE: PRT  
 ORGANISM: Enterococcus faecalis  
 US-09-134-000C-5130  
 Query Match 10.8%; Score 83; DB 4; Length 1216;  
 Best Local Similarity 24.4%; Pred. No. 8.2;  
 Matches 49; Conservative 19; Mismatches 55; Indels 78; Gaps 10;  
 QY 6 VAFAAIVV--SGSALAGVVPQWGG--GGNHNGGSSGPDST-----LSIY- 48  
 DB 181 VTGNTSLVNDSSGRASI---YGGYGNATNTANTVGNVSTKVAITNAATGQFQLSIY 237  
 QY 49 ---QYGSANAALQSDARKSETTITQSGYNGADYVQ-----GADNSTIELTQ 94  
 DB 238 GGQYQYGNIG-----GKVTNDISGYRWYTAGQRFIGGSSRGDICTNRATDGI 286  
 QY 95 N-----GFRNNATYDQLVTRVTHMAHAG-----GNNALVNQTA 130  
 DB 287 NLNTQLYSAGRADFEQNGYSGTIIIGDITNVVATGNSAGGINDFNGGAGNNVSKNS- 345  
 QY 131 SDSSVMVRQVFGNNATANQY 151  
 DB 346 -----QIGASNEATYDAY 358

RESULT 14  
 US-09-489-039A-7849  
 ; Sequence 7849, Application US/09489039A  
 ; Patent No. 6610836  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Gary Breton et. al  
 ; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO KLEBSIELLA  
 ; TITLE OF INVENTION: PNEUMONIAE FOR DIAGNOSTICS AND THERAPEUTICS  
 ; FILE REFERENCE: 2709.2004001  
 ; CURRENT APPLICATION NUMBER: US/09/489,039A  
 ; CURRENT FILING DATE: 2000-01-27  
 ; PRIOR APPLICATION NUMBER: US 60/117,747  
 ; PRIOR FILING DATE: 1999-01-29  
 ; NUMBER OF SEQ ID NOS: 14342  
 ; SEQ ID NO 7849  
 ; LENGTH: 589  
 ; TYPE: PRT  
 ; ORGANISM: Klebsiella pneumoniae  
 US-09-489-039A-7849

Query Match 10.6%; Score 81.5; DB 4; Length 589;  
 Best Local Similarity 24.5%; Pred. No. 4.5;  
 Matches 35; Conservative 21; Mismatches 56; Indels 31; Gaps 7;  
 QY 29 GGNHNGGSSGPDSTLSIYQYGS-----ANAALQSDARKSETTITQSGYNGA- 79  
 DB 259 GGLDRNGANAGQDTPFIYAFDTLITERIEINGGLRDNVHTKYDSATACGSGRAI 318  
 QY 80 --DVQY--GADNSTIELTQNG-----FNNATYDQLVTRVTHMAHAGN---NALVNQ 128  
 DB 319 ACPGQSTGSPVTTVDTKAGNLVNMKAGALYR-----LTEQ-----GNVYVYAISSQ 367  
 QY 129 TASDSSVMVRQVFGNNATANQY 151  
 DB 368 PGGSSFALAASGSGNSANRTDF 390

RESULT 15  
 US-09-540-236-2676  
 ; Sequence 2676, Application US/09540236  
 ; Patent No. 6673910  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Gary L. Breton et al.  
 ; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO MORAXELLA CATARI  
 ; TITLE OF INVENTION: FOR DIAGNOSTICS AND THERAPEUTICS  
 ; FILE REFERENCE: 2709.2005-001  
 ; CURRENT APPLICATION NUMBER: US/09/540,236  
 ; CURRENT FILING DATE: 2000-04-04

```

; NUMBER OF SEQ ID NOS: 3840
; SEQ ID NO 2676
; LENGTH: 867
; TYPE: PRT
; ORGANISM: M.catarrhalis
US-09-540-236-2676

```

Query Match	10.4%;	Score 80;	DB 4;	Length 867;	
Best Local Similarity	26.8%;	Pred. No. 11;			
Matches	49;	Conservative	23;	Mismatches 65;	Indels 46; Gaps 10;
QY	4	LKVAAPAAIIVVSGSALAG--VVPQWGG-----	GGNH-----	GGG-----	NSGQPD 43
DB	33	LGMATTASAOVQAGYIGGNNVNAQTKYSTVSGGDYNEANGELISTIGCGFFNKASGESS	92		
QY	44	TLISIYQGSANAALALOSDARKSETITQSGY-----	GNQADVQCGADNST--	IELTQNGF	97
DB	93	TIA-----GGRN-----	NQATKENSTVGGGKFNQAKGRNSTVAGGYNNEATGIDSTIAGG	142	
QY	98	R--NNATYDQLVTRVVTTEMAHAGNNAAALVNQATSDSSVMV-----	RQVGFQGNNA	147	
DB	143	RGHQATGAGSFAAGVGNGQATGAGSFAAGVGNGQANADNAVAGNKNSTITGKDSVAIGSNNT	202		
QY	148	ANQ	150		
DB	203	VAQ	205		

Search completed: August 2, 2004, 14:58:37  
Job time : 13 secs

This Page Blank (uspto)

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: August 2, 2004, 14:54:48 ; Search time 36.8 Seconds  
(without alignments)  
1287.123 Million cell updates/sec

Title: US-09-543-407-30.

Perfect score: 768

Sequence: 1 MKLLKVAFAAIVVSGSALA.....DSSVMVRQVGFNNATANYQ 151

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1291235 seqs, 313682936 residues

Total number of hits satisfying chosen parameters: 1291235

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

- Published Applications AA:\*
- 1: /cgn2\_6/ptodata/2/pubpaa/US07\_PUBCOMB.pep:\*
  - 2: /cgn2\_6/ptodata/2/pubpaa/PTCT\_NEW\_PUB.pep:\*
  - 3: /cgn2\_6/ptodata/2/pubpaa/US05\_NEW\_PUB.pep:\*
  - 4: /cgn2\_6/ptodata/2/pubpaa/US06\_PUBCOMB.pep:\*
  - 5: /cgn2\_6/ptodata/2/pubpaa/US07\_NEW\_PUB.pep:\*
  - 6: /cgn2\_6/ptodata/2/pubpaa/PTCTUS\_PUBCOMB.pep:\*
  - 7: /cgn2\_6/ptodata/2/pubpaa/US08\_NEW\_PUB.pep:\*
  - 8: /cgn2\_6/ptodata/2/pubpaa/US08\_PUBCOMB.pep:\*
  - 9: /cgn2\_6/ptodata/2/pubpaa/US09A\_PUBCOMB.pep:\*
  - 10: /cgn2\_6/ptodata/2/pubpaa/US09B\_PUBCOMB.pep:\*
  - 11: /cgn2\_6/ptodata/2/pubpaa/US09C\_PUBCOMB.pep:\*
  - 12: /cgn2\_6/ptodata/2/pubpaa/US09D\_PUBCOMB.pep:\*
  - 13: /cgn2\_6/ptodata/2/pubpaa/US10A\_PUBCOMB.pep:\*
  - 14: /cgn2\_6/ptodata/2/pubpaa/US10B\_PUBCOMB.pep:\*
  - 15: /cgn2\_6/ptodata/2/pubpaa/US10C\_PUBCOMB.pep:\*
  - 16: /cgn2\_6/ptodata/2/pubpaa/US10\_NEW\_PUB.pep:\*
  - 17: /cgn2\_6/ptodata/2/pubpaa/US60\_NEW\_PUB.pep:\*
  - 18: /cgn2\_6/ptodata/2/pubpaa/US60\_PUBCOMB.pep:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	51.4	66.9	151	12	US-09-741-873B-4
2	51.4	66.9	151	12	US-09-741-873B-4
3	43.6	56.8	131	12	US-09-741-873B-2
4	43.6	56.8	131	12	US-09-741-873B-2
5	103	13.4	445	15	US-10-363-493-20638
6	94.5	12.3	537	16	US-10-437-963-141342
7	94	12.2	354	10	US-09-820-843A-21
8	94	12.2	498	12	US-10-424-599-196154
9	93.5	12.1	2204	12	US-10-282-122A-64364
10	93	12.1	263	12	US-10-425-114-43960
11	92.5	12.0	1721	12	US-10-282-122A-62548
12	91.5	11.9	673	9	US-09-841-835-8
13	91.5	11.9	949	9	US-09-841-835-10
14	91.5	11.9	1327	9	US-09-841-835-2
15	91.5	11.9	1327	10	US-09-972-115A-8

16	91.5	11.9	1327	14	US-10-199-937-4	Sequence 4, Appli
17	90.5	11.8	628	12	US-10-282-122A-53269	Sequence 53269, A
18	90	11.7	171	12	US-10-424-599-173860	Sequence 173860, A
19	90	11.7	703	16	US-10-437-963-108981	Sequence 108981, A
20	89.5	11.7	1129	12	US-10-282-122A-48048	Sequence 48048, A
21	88.5	11.5	145	16	US-10-437-963-147748	Sequence 147748, A
22	88.5	11.5	1217	14	US-10-311-406-2	Sequence 2, Appli
23	88.5	11.5	1862	12	US-10-282-122A-49757	Sequence 49757, A
24	87.5	11.4	191	16	US-10-437-963-105413	Sequence 105413, A
25	87	11.3	369	12	US-10-425-114-56041	Sequence 56041, A
26	87	11.3	486	12	US-10-424-599-275468	Sequence 275468, A
27	87	11.3	507	12	US-10-425-114-57763	Sequence 57763, A
28	87	11.3	597	9	US-09-793-306-146	Sequence 146, App
29	86.5	11.3	197	12	US-10-425-114-67750	Sequence 67750, A
30	86.5	11.3	745	8	US-08-834-666A-6	Sequence 6, Appli
31	86.5	11.3	745	12	US-10-282-122A-58683	Sequence 58683, A
32	86	11.2	538	12	US-10-425-114-68152	Sequence 68152, A
33	85.5	11.1	270	16	US-10-437-963-122263	Sequence 122263, A
34	85.5	11.1	298	12	US-10-282-122A-47486	Sequence 47486, A
35	85	11.1	321	12	US-10-412-699B-1074	Sequence 1074, Ap
36	85	11.1	321	15	US-10-374-780A-530	Sequence 530, App
37	85	11.1	321	16	US-10-437-963-17276	Sequence 17276, A
38	85	11.1	323	12	US-10-659-869-26	Sequence 26, Appl
39	85	11.1	323	14	US-10-021-811-26	Sequence 26, Appl
40	85	11.1	1778	14	US-10-238-075-749	Sequence 749, App
41	84.5	11.0	322	9	US-09-905-176-24	Sequence 24, Appl
42	84.5	11.0	678	12	US-10-282-122A-64573	Sequence 64573, A
43	84	10.9	65	9	US-09-996-194-16	Sequence 16, Appl
44	84	10.9	65	12	US-10-164-966-33	Sequence 33, Appl
45	84	10.9	478	16	US-10-437-963-115033	Sequence 115033, A

ALIGNMENTS

RESULT 1  
US-09-741-873B-4  
; Sequence 4, Application US/09741873B  
; Publication No. US20020081722A1  
; GENERAL INFORMATION:  
; APPLICANT: Normark, Staffan  
; APPLICANT: Olsen, Arne  
; TITLE OF INVENTION: Fibronectin Binding Protein As Well As Its Preparation  
; FILE REFERENCE: 012889-084  
; CURRENT APPLICATION NUMBER: US/09/741.873B  
; CURRENT FILING DATE: 2003-04-04  
; PRIOR APPLICATION NUMBER: SE 8801723-1  
; PRIOR FILING DATE: 1998-05-06  
; PRIOR APPLICATION NUMBER: US 08/978,878  
; PRIOR FILING DATE: 1997-11-26  
; PRIOR APPLICATION NUMBER: US 07/347,189  
; PRIOR FILING DATE: 1989-05-04  
; PRIOR APPLICATION NUMBER: US 07/789,437  
; PRIOR FILING DATE: 1991-11-06  
; PRIOR APPLICATION NUMBER: US 07/970,846  
; PRIOR FILING DATE: 1992-11-03  
; PRIOR APPLICATION NUMBER: US 08/187,865  
; PRIOR FILING DATE: 1994-01-28  
; PRIOR APPLICATION NUMBER: US 08/318,519  
; PRIOR FILING DATE: 1994-10-05  
; NUMBER OF SEQ ID NOS: 10  
; SOFTWARE: PatentIn version 3.0  
; SEQ ID NO 4  
; LENGTH: 151  
; TYPE: PRT  
; ORGANISM: Escherichia coli  
US-09-741-873B-4

Query Match 66.9%; Score 51.4; DB 12; Length 151;  
Best Local Similarity 66.9%; Pred. No. 1.4e-43;  
Matches 104; Conservative 18; Mismatches 29; Indels 0; Gaps 0;  
QY 1 MKLLKVAFAAIVVSGSALGVVPQWGGGNGHNGSGPDSLTSTIYQYGSANAALQ 60

Db 1 MKLLKVAALVFGSSAVGVFPYGGGNGHGGGNSGPNSELNIYQYGGNSALALQ 60  
 QY 61 SDARKSETTITQSGYNGADVGGQADNSTIELTQNGFRNNATYDQLVTRVVTHEMAHAGG 120  
 Db 61 TDARNSDLTITQGGGNGADVGGQSDSSIDLTRQFGNSATLDQWNGKNSMTVKQFGG 120  
 QY 121 NNAALVNOTASDSSVVMVQVGFNNATANQY 151  
 Db 121 GNGAAVDQTASNSNVNVTQVGFNNATAHQY 151

## RESULT 2

US-09-741-873B-4  
 ; Sequence 4, Application US/09741873B  
 ; Publication No. US20040096965A9  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Normark, Staffan  
 ; APPLICANT: Olsen, Arne  
 ; TITLE OF INVENTION: Fibronectin Binding Protein As Well As Its Preparation  
 ; FILE REFERENCE: 012889-084  
 ; CURRENT APPLICATION NUMBER: US/09/741,873B  
 ; CURRENT FILING DATE: 2003-04-04  
 ; PRIOR APPLICATION NUMBER: SE 8801723-1  
 ; PRIOR FILING DATE: 1998-05-06  
 ; PRIOR APPLICATION NUMBER: US 08/978,878  
 ; PRIOR FILING DATE: 1997-11-26  
 ; PRIOR APPLICATION NUMBER: US 07/347,189  
 ; PRIOR FILING DATE: 1989-05-04  
 ; PRIOR APPLICATION NUMBER: US 07/789,437  
 ; PRIOR FILING DATE: 1991-11-06  
 ; PRIOR APPLICATION NUMBER: US 07/970,846  
 ; PRIOR FILING DATE: 1992-11-03  
 ; PRIOR APPLICATION NUMBER: US 08/187,865  
 ; PRIOR FILING DATE: 1994-01-28  
 ; PRIOR APPLICATION NUMBER: US 08/318,519  
 ; PRIOR FILING DATE: 1994-10-05  
 ; NUMBER OF SEQ ID NOS: 10  
 ; SOFTWARE: Patent in version 3.0  
 ; SEQ ID NO 4  
 ; LENGTH: 151  
 ; TYPE: PRT  
 ; ORGANISM: Escherichia coli  
 ; US-09-741-873B-4

Query Match 66.9%; Score 514; DB 12; Length 151;  
 Best Local Similarity 68.9%; Pred. No. 1.4e-43;  
 Matches 104; Conservative 18; Mismatches 29; Indels 0; Gaps 0;  
 QY 1 MKLLKVAALVFGSSALAGVFPYGGGNGHGGGNSGPNSELNIYQYGSANAALALQ 60  
 Db 1 MKLLKVAALVFGSSAVGVFPYGGGNGHGGGNSGPNSELNIYQYGGNSALALQ 60  
 QY 61 SDARKSETTITQSGYNGADVGGQADNSTIELTQNGFRNNATYDQLVTRVVTHEMAHAGG 120  
 Db 61 TDARNSDLTITQGGGNGADVGGQSDSSIDLTRQFGNSATLDQWNGKNSMTVKQFGG 120  
 QY 121 NNAALVNOTASDSSVVMVQVGFNNATANQY 151  
 Db 121 GNGAAVDQTASNSNVNVTQVGFNNATAHQY 151

## RESULT 3

US-09-741-873B-2  
 ; Sequence 2, Application US/09741873B  
 ; Publication No. US20020081722A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Normark, Staffan  
 ; APPLICANT: Olsen, Arne  
 ; TITLE OF INVENTION: Fibronectin Binding Protein As Well As Its Preparation  
 ; FILE REFERENCE: 012889-084  
 ; CURRENT APPLICATION NUMBER: US/09/741,873B  
 ; CURRENT FILING DATE: 2003-04-04

; PRIOR APPLICATION NUMBER: SE 8801723-1  
 ; PRIOR FILING DATE: 1998-05-06  
 ; PRIOR APPLICATION NUMBER: US 08/978,878  
 ; PRIOR FILING DATE: 1997-11-26  
 ; PRIOR APPLICATION NUMBER: US 07/347,189  
 ; PRIOR FILING DATE: 1989-05-04  
 ; PRIOR APPLICATION NUMBER: US 07/789,437  
 ; PRIOR FILING DATE: 1991-11-06  
 ; PRIOR APPLICATION NUMBER: US 07/970,846  
 ; PRIOR FILING DATE: 1992-11-03  
 ; PRIOR APPLICATION NUMBER: US 08/187,865  
 ; PRIOR FILING DATE: 1994-01-28  
 ; PRIOR APPLICATION NUMBER: US 08/318,519  
 ; PRIOR FILING DATE: 1994-10-05  
 ; NUMBER OF SEQ ID NOS: 10  
 ; SOFTWARE: Patent in version 3.0  
 ; SEQ ID NO 2  
 ; LENGTH: 131  
 ; TYPE: PRT  
 ; ORGANISM: Escherichia coli  
 ; US-09-741-873B-2

Query Match 56.8%; Score 436; DB 12; Length 131;  
 Best Local Similarity 65.6%; Pred. No. 7.7e-36;  
 Matches 86; Conservative 17; Mismatches 28; Indels 0; Gaps 0;  
 QY 21 GVVPQWGGGNGHGGGNSGPNSELNIYQYGSANAALALOSDARKSETTITQSGYNGAD 80  
 Db 1 GVVPQYGGGNGHGGGNSGPNSELNIYQYGGNSALALQTDARNSDLTITQGGGNGAD 60  
 QY 81 VGGADNSTIELTQNGFRNNATYDQLVTRVVTHEMAHAGGNSAALVNOTASDSSVVMVQV 140  
 Db 61 VGGSDSSIDLTRQFGNSATLDQWNGKNSMTVKQFGGNGAAVDQTASNSNVNVTQV 120  
 QY 141 GFGNNATANQY 151  
 Db 121 GFGNNATAHQY 131

## RESULT 4

US-09-741-873B-2  
 ; Sequence 2, Application US/09741873B  
 ; Publication No. US20040096965A9  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Normark, Staffan  
 ; APPLICANT: Olsen, Arne  
 ; TITLE OF INVENTION: Fibronectin Binding Protein As Well As Its Preparation  
 ; FILE REFERENCE: 012889-084  
 ; CURRENT APPLICATION NUMBER: US/09/741,873B  
 ; CURRENT FILING DATE: 2003-04-04  
 ; PRIOR APPLICATION NUMBER: SE 8801723-1  
 ; PRIOR FILING DATE: 1998-05-06  
 ; PRIOR APPLICATION NUMBER: US 08/978,878  
 ; PRIOR FILING DATE: 1997-11-26  
 ; PRIOR APPLICATION NUMBER: US 07/347,189  
 ; PRIOR FILING DATE: 1989-05-04  
 ; PRIOR APPLICATION NUMBER: US 07/789,437  
 ; PRIOR FILING DATE: 1991-11-06  
 ; PRIOR APPLICATION NUMBER: US 07/970,846  
 ; PRIOR FILING DATE: 1992-11-03  
 ; PRIOR APPLICATION NUMBER: US 08/187,865  
 ; PRIOR FILING DATE: 1994-01-28  
 ; PRIOR APPLICATION NUMBER: US 08/318,519  
 ; PRIOR FILING DATE: 1994-10-05  
 ; NUMBER OF SEQ ID NOS: 10  
 ; SOFTWARE: Patent in version 3.0  
 ; SEQ ID NO 2  
 ; LENGTH: 131  
 ; TYPE: PRT  
 ; ORGANISM: Escherichia coli  
 ; US-09-741-873B-2

Query Match

56.8%; Score 436; DB 12; Length 131;



Qy	26	WGCGNHNGGNSGGPDS---	TLSTIYVGSAALALQSDARKSETTI-----	TOSGY. 75
		:	:	
Db	148	WNGGSTNTGLANAGAGNTGFFDA	GNYNFGSLNAGNINSSFCNGSGDGNFLNAGDVNSGV	207
		:	:	
Qy	76	GNAGDVQCGADNSTIELTQNGFN	NATYDQLVTRVVTHEMAHAG-----	GNNAAALNQYT 129
		:	:	
Db	208	GNAGDVNTGLNSG-NINTGGF-N	PGTLNTGFFSAMTQAGPNSGFFNAGTNGSGFGHNDP	265
		:	:	
Qy	130	ASDSSVWVRQVGFGNNTAN		149
		:	:	
Db	266	AGSGNSGIGNSGFGNSGYVN		285
		:	:	

```
RESULT 8
US-10-424-599-196154
; Sequence 196154, Application US/10424599
; Publication No. US20040031072A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa Thomas J
; APPLICANT: Kovalic David K
; APPLICANT: Zhou Yihua
; APPLICANT: Cao Yongwei
; TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated with
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(53223)B
; CURRENT APPLICATION NUMBER: US/10/424,599
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 285684
; SEQ ID NO 196154
; LENGTH: 498
; TYPE: PRT
; ORGANISM: Glycine max
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT3847_19152C.1.pep
US-10-424-599-196154

Query Match 12.2%; Score 94; DB 12; Length 498;
Best Local Similarity 28.1%; Pred. No. 0.72;
Matches 34; Conservative 24; Mismatches 41; Indels 22; Gaps 6;

QY 30 GNHGGGNSGPD-STLSIYQGSANAALALQSDARKSETTITQSGYNG-----A 79
Db 355 GSSNSDNASQDTSBQNNKEGSSDNGS--NDANQNGSNSNENANDNGNASHEAQNA 412
QY 80 DVGGADNSTIEL-----TQGFRRNATYDQLVTRVTHMAHAGG-NNAALVNOTASDSS 134
Db 413 DAGGQNEATVESKASNEGAQNETVESQ-----KEESASHSGGSDNSNLNDQSSDHS 467

QY 135 V 135
Db 468 I 468

RESULT 9
US-10-282-122A-64364
; Sequence 64364, Application US/10282122A
; Publication No. US20040029129A1
; GENERAL INFORMATION:
; APPLICANT: Wang, Liangsu
; APPLICANT: Zamudio, Carlos
; APPLICANT: Malone, Cheryl
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari
; APPLICANT: Zyskind, Judith
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John
; APPLICANT: Carr, Grant
; APPLICANT: Yamamoto, Robert
; APPLICANT: Forsyth, R.
; APPLICANT: Xu, H.
; TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
; FILE REFERENCE: ELITRA.034A
; CURRENT APPLICATION NUMBER: US/10/282,122A
; CURRENT FILING DATE: 2003-02-20
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/230,335
; PRIOR FILING DATE: 2000-09-06
; PRIOR APPLICATION NUMBER: 60/230,347
; PRIOR FILING DATE: 2000-09-09
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
```

```
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/267,636
; PRIOR FILING DATE: 2001-02-09
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 78614
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 64364
; LENGTH: 2204
; TYPE: PRT
; ORGANISM: Mycobacterium tuberculosis
US-10-282-122A-64364

Query Match 12.2%; Score 93.5; DB 12; Length 2204;
Best Local Similarity 26.1%; Pred. No. 5.1;
Matches 43; Conservative 18; Mismatches 63; Indels 41; Gaps 6;

QY 10 AATVVG-----SALAGV-----VPOWGGGHHNGGNSGPDSTLSIYQVGA 53
Db 840 AATVGGPTTSIGITASAGISITIPIDIPATSGFN-----STTSBSSGFFNSGASA 894
QY 54 NAALALQSDARKSETTITQSGYNGADVQGGADN-----STIELTQNGFRNNAT 102
Db 895 SGFLNVVAGASGI-----SGFLNVGALGSGVTNVGHTVSGFYNASALDLVTPAFASGLM 948
QY 103 YDQLVTRVTHMAHAGGNNAALVNOTASDSSVM---VRQVGFEN 144
Db 949 RDGKGTMTNLGLANLGSNNAGFGTGTGFDVGANLGNVNIQFEN 993

RESULT 10
US-10-425-114-49960
; Sequence 49960, Application US/10425114
; Publication No. US20040034888A1
; GENERAL INFORMATION:
; APPLICANT: Liu, Jingdong
; APPLICANT: Zhou, Yihua
; APPLICANT: Kovalic, David K.
; APPLICANT: Screen, Steven E
; APPLICANT: Tabaska, Jack E
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated with
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(53313)B
; CURRENT APPLICATION NUMBER: US/10/425,114
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 73128
; SEQ ID NO 49960
; LENGTH: 263
; TYPE: PRT
; ORGANISM: Zea mays
; FEATURE:
; OTHER INFORMATION: Clone ID: 700071884_FLI.pep
US-10-425-114-49960

Query Match 12.1%; Score 93; DB 12; Length 263;
Best Local Similarity 26.8%; Pred. No. 0.41;
Matches 33; Conservative 16; Mismatches 42; Indels 32; Gaps 5;

QY 17 SALAGVVPOWGGGHHNGGNSGPDSTLSIYQGSANAALALQSDARKSETTITQSG-- 74
Db 84 SSAG-----GGGGGGGGGGTNGGGGGGGGGSSSTAASGPGSSGNVADAEKGAG 138
QY 75 -----YNGGA--DVQGGADNSTIEL--TQGFRRNATYDQLVTRVTHMAHAGG 120
Db 139 GGMGGGANGAYSGAGGGVKGEGVSGVALAPSSNGYNGGAAD-----ATGGG 187
QY 121 NNA 123
```

Db 188 SGA 190

RESULT 11

US-10-282-122A-62548

Sequence 62548, Application US/10282122A

Publication No. US20040029129A1

GENERAL INFORMATION:

APPLICANT: Wang, Liangsu

APPLICANT: Zamudio, Carlos

APPLICANT: Malone, Cheryl

APPLICANT: Haselbeck, Robert

APPLICANT: Ohlsen, Kari

APPLICANT: Zyskind, Judith

APPLICANT: Wall, Daniel

APPLICANT: Trawick, John

APPLICANT: Carr, Grant

APPLICANT: Yamamoto, Robert

APPLICANT: Forsyth, R.

APPLICANT: Xu, H.

TITLE OF INVENTION: Identification of Essential Genes in Microorganisms

FILE REFERENCE: ELITRA.034A

CURRENT APPLICATION NUMBER: US/10/282,122A

CURRENT FILING DATE: 2003-02-20

PRIOR FILING DATE: 2000-03-21

PRIOR APPLICATION NUMBER: 60/191,078

PRIOR FILING DATE: 2000-03-21

PRIOR APPLICATION NUMBER: 60/206,848

PRIOR FILING DATE: 2000-05-23

PRIOR APPLICATION NUMBER: 60/207,727

PRIOR FILING DATE: 2000-05-26

PRIOR APPLICATION NUMBER: 60/230,335

PRIOR FILING DATE: 2000-09-06

PRIOR APPLICATION NUMBER: 60/230,347

PRIOR FILING DATE: 2000-09-09

PRIOR APPLICATION NUMBER: 60/242,578

PRIOR FILING DATE: 2000-10-23

PRIOR APPLICATION NUMBER: 60/253,625

PRIOR FILING DATE: 2000-11-27

PRIOR APPLICATION NUMBER: 60/257,931

PRIOR FILING DATE: 2000-12-22

PRIOR APPLICATION NUMBER: 60/267,636

PRIOR FILING DATE: 2001-02-09

PRIOR APPLICATION NUMBER: 60/269,308

PRIOR FILING DATE: 2001-02-16

Remaining Prior Application data removed - See File Wrapper or PALM.

NUMBER OF SEQ ID NOS: 78614

SOFTWARE: Patent in version 3.1

SEQ ID NO 62548

LENGTH: 1721

TYPE: PRT

ORGANISM: Mycobacterium bovis

US-10-282-122A-62548

Query Match 12.0%; Score 92.5; DB 12; Length 1721;

Best Local Similarity 29.5%; Pred. No. 4.8;

Matches 39; Conservative 18; Mismatches 58; Indels 17; Gaps 6;

QY 23 VPQGGGGNHGNGNS----SGPDSLSTLYQY-GSANAALQSDARKSETTITQSGYN 77

Db 921 IPATPGFNSTGGLSGFNSGASGNGFNGASGFNNLVS-----TTSGMSGFLN 974

QY 78 GADVGGGADNSTLTQNGFRNNATYDQLVTRVWTHMAHAGGNNALVNQTSDDSSVMV 137

Db 975 VGALGSGVAN--VGNTISGIYNGVTS-D-LSTPAVNSGLANIGTNIAGLLRDGAGTAAI-- 1029

QY 138 RQVGFNGNATAN 149

Db 1030 -NLGLAHGNLN 1040

RESULT 12

US-09-841-835-8

Sequence 8, Application US/09841835

GENERAL INFORMATION:

APPLICANT: de Lange, Titia

APPLICANT: Smith, Susan

TITLE OF INVENTION: A PROTEIN THAT BINDS TO TRF1 AND METHODS

NUMBER OF SEQUENCES: 12

CORRESPONDENCE ADDRESS:

ADDRESSEE: Klauber & Jackson

STREET: 411 Hackensack Avenue, 4th Floor

CITY: Hackensack

STATE: New Jersey

COUNTRY: USA

ZIP: 07601

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patent in Release #1.0, Version #1.30

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/09/841,835

FILING DATE:

CLASSIFICATION:

PRIOR APPLICATION DATA:

APPLICATION NUMBER: 09/196,387

FILING DATE:

ATTORNEY/AGENT INFORMATION:

NAME: Jackson Esq., David A.

REGISTRATION NUMBER: 26,742

REFERENCE/DOCKET NUMBER: 600-1-230 CIP1

TELECOMMUNICATION INFORMATION:

TELEPHONE: 201-487-5800

TELEFAX: 201-343-1684

TELEX: 133521

INFORMATION FOR SEQ ID NO: 8:

SEQUENCE CHARACTERISTICS:

LENGTH: 673 amino acids

TYPE: amino acid

TOPOLOGY: linear

MOLECULE TYPE: protein

US-09-841-835-8

Query Match 11.9%; Score 91.5; DB 9; Length 673;

Best Local Similarity 29.4%; Pred. No. 1.9;

Matches 42; Conservative 17; Mismatches 65; Indels 19; Gaps 6;

QY 6 VAAFAAI-VVSGSALAGVVPQWGGGNGHNGGNSGPDSTLSYQYGSANAALALQSDAR 64

Db 99 VAAAPVVPVAVSTSSAAGVAPNPAGSGNNSPSSSPTSS--SSSSPSSPGSSLAESPEAA 157

QY 65 KSETTIT----QSCYGNAGDVGGQADNSTLTQNGFRNNATYDQLVTRVWTHMAHAGG 120

Db 158 GVSSTAPLPGCAAGPGTGVPVAVSGALRELLACRNGD-----VSRV--KRLVDRAAN 206

QY 121 NNAALVNQTSDDSSVMVRQVGF 143

Db 207 VNAK--DMAGKSSPLHFAAGFG 227

RESULT 13

US-09-841-835-10

Sequence 10, Application US/09841835

Patent No. US20020076795A1

GENERAL INFORMATION:

APPLICANT: de Lange, Titia

APPLICANT: Smith, Susan

TITLE OF INVENTION: A PROTEIN THAT BINDS TO TRF1 AND METHODS

NUMBER OF SEQUENCES: 12

CORRESPONDENCE ADDRESS:

ADDRESSEE: Klauber & Jackson

STREET: 411 Hackensack Avenue, 4th Floor

CITY: Hackensack

Patent No. US20020076795A1

GENERAL INFORMATION:

APPLICANT: de Lange, Titia

APPLICANT: Smith, Susan

TITLE OF INVENTION: A PROTEIN THAT BINDS TO TRF1 AND METHODS

NUMBER OF SEQUENCES: 12

CORRESPONDENCE ADDRESS:

ADDRESSEE: Klauber & Jackson

STREET: 411 Hackensack Avenue, 4th Floor

CITY: Hackensack

STATE: New Jersey

COUNTRY: USA

ZIP: 07601

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patent in Release #1.0, Version #1.30

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/09/841,835

FILING DATE:

CLASSIFICATION:

PRIOR APPLICATION DATA:

APPLICATION NUMBER: 09/196,387

FILING DATE:

ATTORNEY/AGENT INFORMATION:

NAME: Jackson Esq., David A.

REGISTRATION NUMBER: 26,742

REFERENCE/DOCKET NUMBER: 600-1-230 CIP1

TELECOMMUNICATION INFORMATION:

TELEPHONE: 201-487-5800

TELEFAX: 201-343-1684

TELEX: 133521

INFORMATION FOR SEQ ID NO: 8:

SEQUENCE CHARACTERISTICS:

LENGTH: 673 amino acids

TYPE: amino acid

TOPOLOGY: linear

MOLECULE TYPE: protein

US-09-841-835-8

Query Match 11.9%; Score 91.5; DB 9; Length 673;

Best Local Similarity 29.4%; Pred. No. 1.9;

Matches 42; Conservative 17; Mismatches 65; Indels 19; Gaps 6;

QY 6 VAAFAAI-VVSGSALAGVVPQWGGGNGHNGGNSGPDSTLSYQYGSANAALALQSDAR 64

Db 99 VAAAPVVPVAVSTSSAAGVAPNPAGSGNNSPSSSPTSS--SSSSPSSPGSSLAESPEAA 157

QY 65 KSETTIT----QSCYGNAGDVGGQADNSTLTQNGFRNNATYDQLVTRVWTHMAHAGG 120

Db 158 GVSSTAPLPGCAAGPGTGVPVAVSGALRELLACRNGD-----VSRV--KRLVDRAAN 206

QY 121 NNAALVNQTSDDSSVMVRQVGF 143

Db 207 VNAK--DMAGKSSPLHFAAGFG 227

RESULT 13

US-09-841-835-10

Sequence 10, Application US/09841835

Patent No. US20020076795A1

GENERAL INFORMATION:

APPLICANT: de Lange, Titia

APPLICANT: Smith, Susan

TITLE OF INVENTION: A PROTEIN THAT BINDS TO TRF1 AND METHODS

NUMBER OF SEQUENCES: 12

CORRESPONDENCE ADDRESS:

ADDRESSEE: Klauber & Jackson

STREET: 411 Hackensack Avenue, 4th Floor

CITY: Hackensack

STATE: New Jersey  
COUNTRY: USA  
ZIP: 07601  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/841.835  
FILING DATE:

CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 09/196,387  
FILING DATE:

ATTORNEY/AGENT INFORMATION:  
NAME: Jackson Esq., David A.  
REGISTRATION NUMBER: 26,742  
REFERENCE/DOCKET NUMBER: 600-1-230 CIP1  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 201-487-5800  
TELEFAX: 201-343-1684  
TELEX: 133521

INFORMATION FOR SEQ ID NO: 10:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 949 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-09-841-835-10

Query Match 11.9%; Score 91.5; DB 9; Length 949;  
Best Local Similarity 29.4%; Pred. No. 2.9;  
Matches 42; Conservative 17; Mismatches 65; Indels 19; Gaps 6;  
QY 6 VAAFAAI-VVSGSALAGVVPQWGGGNGGNGGNSGPDSTLSIYQYGSANAALALQSDAR 64  
DB 99 VAAAPVVPVAVSTSSAAGVAPNPAGSGNNSPSSSSPTSS--SSSSPSPGSSLAESPEAA 157  
QY 65 KSETTIT---QSGYGNAGDVQGGADNSTIETQNGFRNNATYDQLVTRVVTHEMAHAGG 120  
DB 158 GVSTAPLPGCAAGPGTGVPAVSGALRELLEACRNGD-----VSRV--KRLVDAAN 206  
QY 121 NNAALVNQTSDDSSVMVRQVGF 143  
DB 207 VNAX--DMAGKSSPLHFAAGFG 227

RESULT 14  
US-09-841-835-2  
Sequence 2, Application US/09841835  
Patent No. US20020076795A1  
GENERAL INFORMATION:  
APPLICANT: de Lange, Titia  
APPLICANT: Smith, Susan  
TITLE OF INVENTION: A PROTEIN THAT BINDS TO TRF1 AND METHODS  
TITLE OF INVENTION: OF USE THEREOF  
NUMBER OF SEQUENCES: 12  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Klauber & Jackson  
STREET: 411 Hackensack Avenue, 4th Floor  
CITY: Hackensack  
STATE: New Jersey  
COUNTRY: USA  
ZIP: 07601  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/841,835  
FILING DATE:

CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 09/196,387  
FILING DATE:  
ATTORNEY/AGENT INFORMATION:  
NAME: Jackson Esq., David A.  
REGISTRATION NUMBER: 26,742  
REFERENCE/DOCKET NUMBER: 600-1-230 CIP1  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 201-487-5800  
TELEFAX: 201-343-1684  
TELEX: 133521

INFORMATION FOR SEQ ID NO: 2:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1327 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
HYPOTHETICAL: NO  
US-09-841-835-2

Query Match 11.9%; Score 91.5; DB 9; Length 1327;  
Best Local Similarity 29.4%; Pred. No. 4.3;  
Matches 42; Conservative 17; Mismatches 65; Indels 19; Gaps 6;  
QY 6 VAAFAAI-VVSGSALAGVVPQWGGGNGGNGGNSGPDSTLSIYQYGSANAALALQSDAR 64  
DB 99 VAAAPVVPVAVSTSSAAGVAPNPAGSGNNSPSSSSPTSS--SSSSPSPGSSLAESPEAA 157  
QY 65 KSETTIT---QSGYGNAGDVQGGADNSTIETQNGFRNNATYDQLVTRVVTHEMAHAGG 120  
DB 158 GVSTAPLPGCAAGPGTGVPAVSGALRELLEACRNGD-----VSRV--KRLVDAAN 206  
QY 121 NNAALVNQTSDDSSVMVRQVGF 143  
DB 207 VNAX--DMAGKSSPLHFAAGFG 227

RESULT 15  
US-09-972-115A-8  
Sequence 8, Application US/09972115A  
Publication No. US20030032769A1  
GENERAL INFORMATION:  
APPLICANT: Geron Corporation  
APPLICANT: Gregg, Morin B.  
APPLICANT: Walter, Funk D.  
APPLICANT: Mieczyslaw, Piatysek A.  
TITLE OF INVENTION: A Second Mammalian Telomerase  
FILE REFERENCE: 080/003C  
CURRENT APPLICATION NUMBER: US/09/972,115A  
CURRENT FILING DATE: 2001-10-05  
PRIOR APPLICATION NUMBER: US 60/128,577  
PRIOR FILING DATE: 2000-04-10  
PRIOR APPLICATION NUMBER: US 60/129,123  
PRIOR FILING DATE: 1999-04-13  
NUMBER OF SEQ ID NOS: 64  
SOFTWARE: PatentIn version 3.1  
SEQ ID NO 8  
LENGTH: 1327  
TYPE: PRT  
ORGANISM: Homo sapiens  
US-09-972-115A-8

Query Match 11.9%; Score 91.5; DB 10; Length 1327;  
Best Local Similarity 29.4%; Pred. No. 4.3;  
Matches 42; Conservative 17; Mismatches 65; Indels 19; Gaps 6;  
QY 6 VAAFAAI-VVSGSALAGVVPQWGGGNGGNGGNSGPDSTLSIYQYGSANAALALQSDAR 64  
DB 99 VAAAPVVPVAVSTSSAAGVAPNPAGSGNNSPSSSSPTSS--SSSSPSPGSSLAESPEAA 157  
QY 65 KSETTIT---QSGYGNAGDVQGGADNSTIETQNGFRNNATYDQLVTRVVTHEMAHAGG 120

Db	158	GVSSTAPLGFGAGPCTGVPVSGALRELEACRNGD-----VSRV--KRLVDAAN	206
Qy	121	NNAALVNQJASDSSVMVRQVFG	143
Db	207	VNAK--DMAGRKSSPLHFAAGFG	227

Search completed: August 2, 2004, 15:36:14  
Job time : 37.8 secs

This Page Blank (uspto)

GenCore version 5.1.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: August 2, 2004, 14:48:33 ; Search time 167.9 Seconds  
(without alignments)  
877.809 Million cell updates/sec

Title: US-09-543-407-30

Perfect score: 768

Sequence: 1 MKLLKVAAPAAIVVSGSALA.....DSSVMVRQVGFNNATANYQ 151

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 6019581 seqs, 976053577 residues

Total number of hits satisfying chosen parameters: 6019581

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Pending Patents AA Main:\*

- 1: /cgn2\_6/ptodata/2/paa/pctus COMB.pep.\*
- 2: /cgn2\_6/ptodata/2/paa/US05 COMB.pep.\*
- 3: /cgn2\_6/ptodata/2/paa/US07 COMB.pep.\*
- 4: /cgn2\_6/ptodata/2/paa/US08 COMB.pep.\*
- 5: /cgn2\_6/ptodata/2/paa/US08 COMB.pep.\*
- 6: /cgn2\_6/ptodata/2/paa/US08 COMB.pep.\*
- 7: /cgn2\_6/ptodata/2/paa/US08 COMB.pep.\*
- 8: /cgn2\_6/ptodata/2/paa/US08 COMB.pep.\*
- 9: /cgn2\_6/ptodata/2/paa/US08 COMB.pep.\*
- 10: /cgn2\_6/ptodata/2/paa/US08 COMB.pep.\*
- 11: /cgn2\_6/ptodata/2/paa/US08 COMB.pep.\*
- 12: /cgn2\_6/ptodata/2/paa/US08 COMB.pep.\*
- 13: /cgn2\_6/ptodata/2/paa/US08 COMB.pep.\*
- 14: /cgn2\_6/ptodata/2/paa/US08 COMB.pep.\*
- 15: /cgn2\_6/ptodata/2/paa/US08 COMB.pep.\*
- 16: /cgn2\_6/ptodata/2/paa/US08 COMB.pep.\*
- 17: /cgn2\_6/ptodata/2/paa/US08 COMB.pep.\*
- 18: /cgn2\_6/ptodata/2/paa/US08 COMB.pep.\*
- 19: /cgn2\_6/ptodata/2/paa/US08 COMB.pep.\*
- 20: /cgn2\_6/ptodata/2/paa/US08 COMB.pep.\*
- 21: /cgn2\_6/ptodata/2/paa/US08 COMB.pep.\*
- 22: /cgn2\_6/ptodata/2/paa/US08 COMB.pep.\*
- 23: /cgn2\_6/ptodata/2/paa/US08 COMB.pep.\*
- 24: /cgn2\_6/ptodata/2/paa/US08 COMB.pep.\*
- 25: /cgn2\_6/ptodata/2/paa/US08 COMB.pep.\*
- 26: /cgn2\_6/ptodata/2/paa/US08 COMB.pep.\*
- 27: /cgn2\_6/ptodata/2/paa/US08 COMB.pep.\*
- 28: /cgn2\_6/ptodata/2/paa/US08 COMB.pep.\*
- 29: /cgn2\_6/ptodata/2/paa/US08 COMB.pep.\*
- 30: /cgn2\_6/ptodata/2/paa/US08 COMB.pep.\*
- 31: /cgn2\_6/ptodata/2/paa/US08 COMB.pep.\*
- 32: /cgn2\_6/ptodata/2/paa/US08 COMB.pep.\*
- 33: /cgn2\_6/ptodata/2/paa/US08 COMB.pep.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match Length	DB ID	Description
------------	-------	--------------------	-------	-------------

1	768	100.0	151	19	US-09-543-407-30	Sequence 30, Appl
2	700	91.1	151	19	US-09-543-407-24	Sequence 24, Appl
3	681	88.7	151	19	US-09-543-407-5	Sequence 5, Appl
4	676	88.0	151	6	US-08-233-642A-57	Sequence 57, Appl
5	659	85.8	151	19	US-09-543-407-14	Sequence 14, Appl
6	657	85.5	151	19	US-09-543-407-22	Sequence 22, Appl
7	613	79.8	151	19	US-09-543-407-28	Sequence 28, Appl
8	603	78.5	151	19	US-09-543-407-12	Sequence 12, Appl
9	600	78.1	151	19	US-09-543-407-26	Sequence 26, Appl
10	599	78.0	151	19	US-09-543-407-18	Sequence 18, Appl
11	597	77.7	151	19	US-09-543-407-20	Sequence 20, Appl
12	594	77.3	131	19	US-09-543-407-31	Sequence 31, Appl
13	566	73.7	151	19	US-09-543-407-16	Sequence 16, Appl
14	517	67.3	151	19	US-09-543-407-7	Sequence 7, Appl
15	514	66.9	151	13	US-08-978-878-4	Sequence 4, Appl
16	514	66.9	151	21	US-09-741-873B-4	Sequence 8, Appl
17	512	66.7	151	33	US-60-352-946-2	Sequence 2, Appl
18	512	66.7	151	33	US-60-444-371-2	Sequence 2, Appl
19	496	64.6	120	6	US-08-233-642A-55	Sequence 5, Appl
20	475	61.8	158	16	US-09-252-691-5834	Sequence 5834, Ap
21	475	61.8	158	16	US-09-252-691C-5834	Sequence 5834, Ap
22	465	61.8	158	30	US-10-417-886-5834	Sequence 34, Ap
23	462	60.2	109	19	US-09-543-407-34	Sequence 2, Appl
24	436	56.8	131	13	US-08-978-878-2	Sequence 2, Appl
25	436	56.8	131	21	US-09-741-873B-2	Sequence 35, Appl
26	332	43.2	109	19	US-09-543-407-35	Sequence 37, Appl
27	310	40.4	68	19	US-09-543-407-39	Sequence 39, Appl
28	237	30.9	48	19	US-09-543-407-39	Sequence 32, Appl
29	217.5	28.3	70	19	US-09-543-407-32	Sequence 32, Appl
30	114.5	14.9	186	16	US-09-252-691-5833	Sequence 5833, Ap
31	114.5	14.9	186	16	US-09-252-691C-5833	Sequence 5833, Ap
32	114.5	14.9	186	30	US-10-417-886-5833	Sequence 8, Appl
33	104.5	13.6	151	19	US-09-543-407-8	Sequence 2, Appl
34	103	13.4	445	29	US-10-369-493-20638	Sequence 20638, A
35	103	13.4	445	33	US-60-360-039-20638	Sequence 21, Appl
36	101.5	13.2	520	1	PCT-US02-18256-21	Sequence 6, Appl
37	99.5	13.0	151	19	US-09-543-407-6	Sequence 8854, Ap
38	98	12.8	145	21	US-09-739-449-8854	Sequence 8854, Ap
39	98	12.8	145	23	US-09-803-110-8854	Sequence 22325, A
40	98	12.8	375	23	US-09-733-089-22325	Sequence 22325, A
41	98	12.8	375	23	US-09-816-660-22325	Sequence 23533, A
42	98	12.8	563	21	US-09-733-089-23533	Sequence 23533, A
43	98	12.6	563	23	US-09-816-660-23533	Sequence 9129, Ap
44	96.5	12.6	313	21	US-09-708-427-9129	Sequence 358, App
45	96.5	12.6	1249	30	US-10-455-719-358	

ALIGNMENTS

RESULT 1  
US-09-543-407-30  
; Sequence 30, Application US/09543407  
; GENERAL INFORMATION:  
; APPLICANT: White, Aaron P.  
; APPLICANT: Doran, James L.  
; APPLICANT: Collinson, S. Karen  
; APPLICANT: Kay, William W.  
; TITLE OF INVENTION: BACTERIAL FIMBRIAL SYSTEM FOR  
; PRESENTATION OF HETEROLOGOUS PEPTIDE SEQUENCES  
; FILE REFERENCE: 920043.406  
; CURRENT FILING DATE: 2000-04-05  
; NUMBER OF SEQ ID NOS: 59  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 30  
; LENGTH: 151  
; TYPE: PRT  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: Recombinant Salmonella enteritidis 3b afgA  
; OTHER INFORMATION: sequence containing the replacement fragment  
; OTHER INFORMATION: encoding PT3 from GP63 of Leishmania major.

US-09-543-407-30

Query Match 100.0%; Score 768; DB 19; Length 151;  
Best Local Similarity 100.0%; Pred. No. 7e-73;  
Matches 151; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MKLLKVAFAAIVVSGSALAGVVPQWGGGNGHNGGNSGPDSTLSIYQYGSANAALALQ 60  
DB 1 MKLLKVAFAAIVVSGSALAGVVPQWGGGNGHNGGNSGPDSTLSIYQYGSANAALALQ 60

QY 61 SDARKSETTITQSGYNGADVGGADNSTIELTQNGFRNNATYDQLVTRVVTHEMAHAGG 120  
DB 61 SDARKSETTITQSGYNGADVGGADNSTIELTQNGFRNNATYDQLVTRVVTHEMAHAGG 120

QY 121 NNAALVNQTASDSSVMVRQVGFNNATANQY 151  
DB 121 NNAALVNQTASDSSVMVRQVGFNNATANQY 151

## RESULT 2

US-09-543-407-24  
; Sequence 24, Application US/09543407  
; GENERAL INFORMATION:  
; APPLICANT: White, Aaron P.  
; APPLICANT: Doran, James L.  
; APPLICANT: Collinson, S. Karen  
; APPLICANT: Kay, William W.  
; TITLE OF INVENTION: BACTERIAL FIMBRIAL SYSTEM FOR  
; FILE REFERENCE: PRESENTATION OF HETEROLOGOUS PEPTIDE SEQUENCES  
; CURRENT APPLICATION NUMBER: US/09/543,407  
; CURRENT FILING DATE: 2000-04-05  
; NUMBER OF SEQ ID NOS: 59  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 24  
; LENGTH: 151  
; TYPE: PRT  
; ORGANISM: Artificial Sequence  
; OTHER INFORMATION: Recombinant Salmonella enteritidis 3b afga  
; OTHER INFORMATION: sequence containing the replacement fragment  
; OTHER INFORMATION: encoding P13 from GP63 of Leishmania major.

US-09-543-407-24

Query Match 91.1%; Score 700; DB 19; Length 151;  
Best Local Similarity 89.9%; Pred. No. 1.2e-65;  
Matches 143; Conservative 0; Mismatches 0; Indels 16; Gaps 2;

QY 1 MKLLKVAFAAIVVSGSALAGVVPQWGGGNGHNGGNSGPDSTLSIYQYGSANAALALQ 60  
DB 1 MKLLKVAFAAIVVSGSALAGVVPQWGGGNGHNGGNSGPDSTLSIYQYGSANAALALQ 60

QY 61 SDARKSETTITQSGYNGADVGGADNSTIELTQNGFRNNAT-----YDQLVTRVVT 112  
DB 61 SDARKSETTITQSGYNGADVGGADNSTIELTQNGFRNNATIDQWNAKNDYDQLVTRVVT 120

QY 113 HEMAHAGGNNALVNQTASDSSVMVRQVGFNNATANQY 151  
DB 121 HEMAHAGG-----NQTASDSSVMVRQVGFNNATANQY 151

## RESULT 3

US-09-543-407-5  
; Sequence 5, Application US/09543407  
; GENERAL INFORMATION:  
; APPLICANT: White, Aaron P.  
; APPLICANT: Doran, James L.  
; APPLICANT: Collinson, S. Karen  
; APPLICANT: Kay, William W.  
; TITLE OF INVENTION: BACTERIAL FIMBRIAL SYSTEM FOR  
; FILE REFERENCE: PRESENTATION OF HETEROLOGOUS PEPTIDE SEQUENCES  
; CURRENT APPLICATION NUMBER: US/09/543,407

US-09-543-407-5  
; CURRENT FILING DATE: 2000-04-05  
; NUMBER OF SEQ ID NOS: 59  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 5  
; LENGTH: 151  
; TYPE: PRT  
; ORGANISM: Salmonella enteritidis

US-09-543-407-5

Query Match 88.7%; Score 681; DB 19; Length 151;  
Best Local Similarity 90.7%; Pred. No. 1.2e-63;  
Matches 137; Conservative 2; Mismatches 12; Indels 0; Gaps 0;

QY 1 MKLLKVAFAAIVVSGSALAGVVPQWGGGNGHNGGNSGPDSTLSIYQYGSANAALALQ 60  
DB 1 MKLLKVAFAAIVVSGSALAGVVPQWGGGNGHNGGNSGPDSTLSIYQYGSANAALALQ 60

QY 61 SDARKSETTITQSGYNGADVGGADNSTIELTQNGFRNNATYDQLVTRVVTHEMAHAGG 120  
DB 61 SDARKSETTITQSGYNGADVGGADNSTIELTQNGFRNNATIDQWNAKNDITVQYGG 120

QY 121 NNAALVNQTASDSSVMVRQVGFNNATANQY 151  
DB 121 NNAALVNQTASDSSVMVRQVGFNNATANQY 151

## RESULT 4

US-08-233-642A-57  
; Sequence 57, Application US/08233642A  
; GENERAL INFORMATION:  
; APPLICANT: Kay, William W.  
; APPLICANT: Collinson, S. Karen  
; APPLICANT: Clouthier, Sharon C.  
; APPLICANT: Doran, James L.  
; TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR SALMONELLA-  
; TITLE OF INVENTION: BASED VACCINES  
; NUMBER OF SEQUENCES: 58  
; NUMBER OF SEQUENCES:  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Seed and Berry  
; STREET: 6300 Columbia Center, 701 Fifth Avenue  
; CITY: Seattle  
; STATE: Washington  
; COUNTRY: U.S.A.  
; ZIP: 98104-7092  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/233,642A  
; FILING DATE: 26-APR-1994  
; CLASSIFICATION: 424  
; ATTORNEY/AGENT INFORMATION:  
; NAME: King, Joshua  
; REGISTRATION NUMBER: 35,570  
; REFERENCE/DOCKET NUMBER: 920043.403C3  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (206) 622-4900  
; TELEFAX: (206) 682-6031  
; TELETYPE: 3723836 SEEDANBERRY  
; INFORMATION FOR SEQ ID NO: 57:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 151 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein

US-08-233-642A-57

Query Match 88.0%; Score 676; DB 6; Length 151;  
Best Local Similarity 90.1%; Pred. No. 4.2e-63;  
Matches 136; Conservative 2; Mismatches 13; Indels 0; Gaps 0;



QY 1 MKLLKVAAPAAIVVSGSALAGVVPQWGGGNGHNGGNSGPDSTLSIYQYGSANAALAIQ 60  
DB 1 MKLLKVAAPAAIVVSGSALAGVVPQWGGGNGHNGGNSGPDSTLSIYQYGSANAALAIQ 60  
QY 61 SDARKSETTITQSGYNGADVGQADNSTIETQNGFRNNATYDQLVTRVVTHEMAHAGG 120  
DB 61 SDARKSETTITQSGYNGADVGQADNSTIETQNGFRNNATIDQWNAKNSDITVQYGG 120  
QY 121 NNPAALVNOTASDSSVMVRQVGFGNNTANQY 151  
DB 121 NNPAALVNOTASDSSVMVRQVGFGNNTANQY 151

RESULT 5  
US-09-543-407-14  
; Sequence 14, Application US/09543407  
; GENERAL INFORMATION:  
; APPLICANT: White, Aaron P.  
; APPLICANT: Doran, James L.  
; APPLICANT: Collinson, S. Karen  
; APPLICANT: Kay, William W.  
; TITLE OF INVENTION: BACTERIAL FIMBRIAL SYSTEM FOR  
; TITLE OF INVENTION: PRESENTATION OF HETEROLOGOUS PEPTIDE SEQUENCES  
; FILE REFERENCE: 920043.406  
; CURRENT APPLICATION NUMBER: US/09/543,407  
; CURRENT FILING DATE: 2000-04-05  
; NUMBER OF SEQ ID NOS: 59  
; SOFTWARE: FastSEQ for Windows Version 4.0  
; SEQ ID NO 14  
; LENGTH: 151  
; TYPE: PRT  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: Recombinant Salmonella enteritidis 3b afga  
; OTHER INFORMATION: sequence containing the replacement fragment  
; OTHER INFORMATION: encoding PT3 from GP63 of Leishmania major.  
US-09-543-407-14

Query Match 85.8%; Score 659; DB 19; Length 151;  
Best Local Similarity 81.9%; Pred. No. 2.7e-61;  
Matches 136; Conservative 0; Mismatches 0; Indels 30; Gaps 2;  
QY 1 MKLLKVAAPAAIVVSGSALAGVVPQWGGGNGHNGGNSGPDSTLSIYQYGSANAALAIQ 60  
DB 1 MKLLKVAAPAAIVVSGSALAGVVPQWGGGNGHNGGNSGPDSTLSIYQYGSANAALAIQ 60  
QY 61 SDARKSETTITQSGYNGADVGQADNSTIETQNGFRNNAT-----YDQ 105  
DB 61 SDARKSETTITQSGYNGADVGQADNSTIETQNGFRNNATIDQWNAKNSDITVQYDQ 120  
QY 106 LVTRVVTHEMAHAGGNAALVNQTASDSSVMVRQVGFGNNTANQY 151  
DB 121 LVTRVVTHEMAH-----SVMVRQVGFGNNTANQY 151

RESULT 6  
US-09-543-407-22  
; Sequence 22, Application US/09543407  
; GENERAL INFORMATION:  
; APPLICANT: White, Aaron P.  
; APPLICANT: Doran, James L.  
; APPLICANT: Collinson, S. Karen  
; APPLICANT: Kay, William W.  
; TITLE OF INVENTION: BACTERIAL FIMBRIAL SYSTEM FOR  
; TITLE OF INVENTION: PRESENTATION OF HETEROLOGOUS PEPTIDE SEQUENCES  
; FILE REFERENCE: 920043.406  
; CURRENT APPLICATION NUMBER: US/09/543,407  
; CURRENT FILING DATE: 2000-04-05  
; NUMBER OF SEQ ID NOS: 59  
; SOFTWARE: FastSEQ for Windows Version 4.0  
; SEQ ID NO 22  
; LENGTH: 151

; TYPE: PRT  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: Recombinant Salmonella enteritidis 3b afga  
; OTHER INFORMATION: sequence containing the replacement fragment  
; OTHER INFORMATION: encoding PT3 from GP63 of Leishmania major.  
US-09-543-407-22  
Query Match 85.5%; Score 657; DB 19; Length 151;  
Best Local Similarity 81.9%; Pred. No. 4.3e-61;  
Matches 136; Conservative 0; Mismatches 0; Indels 30; Gaps 2;  
QY 1 MKLLKVAAPAAIVVSGSALAGVVPQWGGGNGHNGGNSGPDSTLSIYQYGSANAALAIQ 60  
DB 1 MKLLKVAAPAAIVVSGSALAGVVPQWGGGNGHNGGNSGPDSTLSIYQYGSANAALAIQ 60  
QY 61 SDARKSETTITQSGYNGADVGQADNSTIETQNGFRNNATYDQLVTRVVTHEMAH-- 118  
DB 61 SDARKSETTITQSGYNGADVGQADN-----YDQLVTRVVTHEMAHADQ 105  
QY 119 -----CGNNAALVNQTASDSSVMVRQVGFGNNTANQY 151  
DB 106 WNAKNSDITVQYGGNNAALVNQTASDSSVMVRQVGFGNNTANQY 151

RESULT 7  
US-09-543-407-28  
; Sequence 28, Application US/09543407  
; GENERAL INFORMATION:  
; APPLICANT: White, Aaron P.  
; APPLICANT: Doran, James L.  
; APPLICANT: Collinson, S. Karen  
; APPLICANT: Kay, William W.  
; TITLE OF INVENTION: BACTERIAL FIMBRIAL SYSTEM FOR  
; TITLE OF INVENTION: PRESENTATION OF HETEROLOGOUS PEPTIDE SEQUENCES  
; FILE REFERENCE: 920043.406  
; CURRENT APPLICATION NUMBER: US/09/543,407  
; CURRENT FILING DATE: 2000-04-05  
; NUMBER OF SEQ ID NOS: 59  
; SOFTWARE: FastSEQ for Windows Version 4.0  
; SEQ ID NO 28  
; LENGTH: 151  
; TYPE: PRT  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: Recombinant Salmonella enteritidis 3b afga  
; OTHER INFORMATION: sequence containing the replacement fragment  
; OTHER INFORMATION: encoding PT3 from GP63 of Leishmania major.  
US-09-543-407-28

Query Match 79.8%; Score 613; DB 19; Length 151;  
Best Local Similarity 74.6%; Pred. No. 2e-56;  
Matches 129; Conservative 0; Mismatches 0; Indels 44; Gaps 2;  
QY 1 MKLLKVAAPAAIVVSGSALAGVVPQWGGGNGHNGGNSGPDSTLSIYQYGSANAALAIQ 60  
DB 1 MKLLKVAAPAAIVVSGSALAGVVPQWGGGNGHNGGNSGPDSTLSIYQYGSANAALAIQ 60  
QY 61 SDARKSETTITQSGYNGADVGQADNSTIETQNGFRNNATYDQLVTRVVTHEMAH-- 118  
DB 61 SDARKSETTITQSGYNGAD-----YDQLVTRVVTHEMAHAFR 98  
QY 119 -----CGNNAALVNQTASDSSVMVRQVGFGNNTANQY 151  
DB 99 NNATIDWNAKNSDITVQYGGNNAALVNQTASDSSVMVRQVGFGNNTANQY 151

RESULT 8  
US-09-543-407-12  
; Sequence 12, Application US/09543407  
; GENERAL INFORMATION:  
; APPLICANT: White, Aaron P.  
; APPLICANT: Doran, James L.

APPLICANT: Collinson, S. Karen  
APPLICANT: Kay, William W.  
TITLE OF INVENTION: BACTERIAL FIMBRIAL SYSTEM FOR  
PRESENTATION OF HETEROLOGOUS PEPTIDE SEQUENCES  
FILE REFERENCE: 920043.406  
CURRENT APPLICATION NUMBER: US/09/543,407  
CURRENT FILING DATE: 2000-04-05  
NUMBER OF SEQ ID NOS: 59  
SOFTWARE: FastSeq for Windows Version 4.0  
SEQ ID NO 12  
LENGTH: 151  
TYPE: PRT  
ORGANISM: Artificial Sequence  
FEATURE:  
OTHER INFORMATION: Recombinant Salmonella enteritidis 3b afga  
OTHER INFORMATION: sequence containing the replacement fragment  
OTHER INFORMATION: encoding PT3 from GP63 of Leishmania major.  
US-09-543-407-12

Query Match 78.5%; Score 603; DB 19; Length 151;  
Best Local Similarity 80.8%; Pred. No. 2.4e-55;  
Matches 122; Conservative 6; Mismatches 23; Indels 0; Gaps 0;  
QY 1 MKLLKVAAPAAIVVSGSALAGVVPQWGGGNGHNGGNSGPDSTLSIYQYGSANALALQ 60  
DB 1 MKLLKVAAPAAIVVSGSALAGVVPQWGGGNGHNGGNSGPDSTLSIYQYGSANALALQ 60  
QY 61 SDARKSETTITQSGYNGADVGCGADNSTIELTQNGFRNNATYDQLVTRVVTHEMAHAGG 120  
DB 61 SDARKSETTITQSGYNGADVGCGADNSTIELTQNGFRNNATYDQLVTRVVTHEMAHAGG 120  
QY 121 NNAALVNOTASDSSVMVRQVGFNNATANQY 151  
DB 121 NNAALVNOTASDSSVMVRQVGFNNATANQY 151

RESULT 9  
US-09-543-407-26  
Sequence 26, Application US/09543407  
GENERAL INFORMATION:  
APPLICANT: White, Aaron P.  
APPLICANT: Doran, James L.  
APPLICANT: Collinson, S. Karen  
TITLE OF INVENTION: BACTERIAL FIMBRIAL SYSTEM FOR  
PRESENTATION OF HETEROLOGOUS PEPTIDE SEQUENCES  
FILE REFERENCE: 920043.406  
CURRENT APPLICATION NUMBER: US/09/543,407  
CURRENT FILING DATE: 2000-04-05  
NUMBER OF SEQ ID NOS: 59  
SOFTWARE: FastSeq for Windows Version 4.0  
SEQ ID NO 26  
LENGTH: 151  
TYPE: PRT  
ORGANISM: Artificial Sequence  
FEATURE:  
OTHER INFORMATION: Recombinant Salmonella enteritidis 3b afga  
OTHER INFORMATION: sequence containing the replacement fragment  
OTHER INFORMATION: encoding PT3 from GP63 of Leishmania major.  
US-09-543-407-26

Query Match 78.1%; Score 600; DB 19; Length 151;  
Best Local Similarity 81.5%; Pred. No. 4.9e-55;  
Matches 123; Conservative 4; Mismatches 24; Indels 0; Gaps 0;  
QY 1 MKLLKVAAPAAIVVSGSALAGVVPQWGGGNGHNGGNSGPDSTLSIYQYGSANALALQ 60  
DB 1 MKLLKVAAPAAIVVSGSALAGVVPQWGGGNGHNGGNSGPDSTLSIYQYGSANALALQ 60  
QY 61 SDARKSETTITQSGYNGADVGCGADNSTIELTQNGFRNNATYDQLVTRVVTHEMAHAGG 120  
DB 61 LVTRVVTHEMAHAGYNGADVGCGADNSTIELTQNGFRNNATYDQLVTRVVTHEMAHAGG 120

QY 121 NNAALVNOTASDSSVMVRQVGFNNATANQY 151  
DB 121 NNAALVNOTASDSSVMVRQVGFNNATANQY 151

RESULT 10  
US-09-543-407-18  
Sequence 18, Application US/09543407  
GENERAL INFORMATION:  
APPLICANT: White, Aaron P.  
APPLICANT: Doran, James L.  
APPLICANT: Collinson, S. Karen  
TITLE OF INVENTION: BACTERIAL FIMBRIAL SYSTEM FOR  
PRESENTATION OF HETEROLOGOUS PEPTIDE SEQUENCES  
FILE REFERENCE: 920043.406  
CURRENT APPLICATION NUMBER: US/09/543,407  
CURRENT FILING DATE: 2000-04-05  
NUMBER OF SEQ ID NOS: 59  
SOFTWARE: FastSeq for Windows Version 4.0  
SEQ ID NO 18  
LENGTH: 151  
TYPE: PRT  
ORGANISM: Artificial Sequence  
FEATURE:  
OTHER INFORMATION: Recombinant Salmonella enteritidis 3b afga  
OTHER INFORMATION: sequence containing the replacement fragment  
OTHER INFORMATION: encoding PT3 from GP63 of Leishmania major.  
US-09-543-407-18

Query Match 78.0%; Score 599; DB 19; Length 151;  
Best Local Similarity 81.5%; Pred. No. 6.3e-55;  
Matches 123; Conservative 5; Mismatches 23; Indels 0; Gaps 0;  
QY 1 MKLLKVAAPAAIVVSGSALAGVVPQWGGGNGHNGGNSGPDSTLSIYQYGSANALALQ 60  
DB 1 MKLLKVAAPAAIVVSGSALAGVVPQWGGGNGHNGGNSGPDSTLSIYQYGSANALALQ 60  
QY 61 SDARKSETTITQSGYNGADVGCGADNSTIELTQNGFRNNATYDQLVTRVVTHEMAHAGG 120  
DB 61 SDARKSETTITQSGYNGADVGCGADNSTIELTQNGFRNNATYDQLVTRVVTHEMAHAGG 120  
QY 121 NNAALVNOTASDSSVMVRQVGFNNATANQY 151  
DB 121 NNAALVNOTASDSSVMVRQVGFNNATANQY 151

RESULT 11  
US-09-543-407-20  
Sequence 20, Application US/09543407  
GENERAL INFORMATION:  
APPLICANT: White, Aaron P.  
APPLICANT: Doran, James L.  
APPLICANT: Collinson, S. Karen  
TITLE OF INVENTION: BACTERIAL FIMBRIAL SYSTEM FOR  
PRESENTATION OF HETEROLOGOUS PEPTIDE SEQUENCES  
FILE REFERENCE: 920043.406  
CURRENT APPLICATION NUMBER: US/09/543,407  
CURRENT FILING DATE: 2000-04-05  
NUMBER OF SEQ ID NOS: 59  
SOFTWARE: FastSeq for Windows Version 4.0  
SEQ ID NO 20  
LENGTH: 151  
TYPE: PRT  
ORGANISM: Artificial Sequence  
FEATURE:  
OTHER INFORMATION: Recombinant Salmonella enteritidis 3b afga  
OTHER INFORMATION: sequence containing the replacement fragment  
OTHER INFORMATION: encoding PT3 from GP63 of Leishmania major.  
US-09-543-407-20

Query Match 77.7%; Score 597; DB 19; Length 151;

```

RESULT 13
US-09-543-407-16
; Sequence 16, Application US/09543407
; GENERAL INFORMATION:
; APPLICANT: White, Aaron P.
; APPLICANT: Doran, James L.
; APPLICANT: Collinson, S. Karen
; APPLICANT: Kay, William W.
; TITLE OF INVENTION: BACTERIAL FIMBRIAL SYSTEM FOR
; PRESENTATION OF HETEROLOGOUS PEPTIDE SEQUENCES
; TITLE OF INVENTION: PRESENTATION OF HETEROLOGOUS
; FILE REFERENCE: 920043.406
; CURRENT APPLICATION NUMBER: US/09/543,407
; CURRENT FILING DATE: 2000-04-05
; NUMBER OF SEQ ID NOS: 59
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 16
; LENGTH: 151
; TYPE: PRT
; ORGANISM: Artificial Sequence

```

RESULT 15  
US-08-978-878-4  
; Sequence 4, Application US/08978878  
; GENERAL INFORMATION:  
; APPLICANT: NORMARK, Staffan  
; APPLICANT: OLSEN, Arne  
; TITLE OF INVENTION: FIBRONECTIN BINDING PROTEIN AS WELL AS ITS PREPARATION  
; FILE REFERENCE: 012889-081  
; CURRENT APPLICATION NUMBER: US/08/978,878  
; CURRENT FILING DATE: 1997-11-26  
; EARLIER APPLICATION NUMBER: SE 8801723-1  
; EARLIER FILING DATE: 1988-05-06

```

; EARLIER APPLICATION NUMBER: US 07/347,189
; EARLIER FILING DATE: 1989-05-04
; EARLIER APPLICATION NUMBER: US 07/789,437
; EARLIER FILING DATE: 1991-11-06
; EARLIER APPLICATION NUMBER: US 07/970,846
; EARLIER FILING DATE: 1992-11-03
; EARLIER APPLICATION NUMBER: US 08/187,865
; EARLIER FILING DATE: 1994-01-28
; EARLIER APPLICATION NUMBER: US 08/318,519
; EARLIER FILING DATE: 1994-10-05
; EARLIER APPLICATION NUMBER: US 08/495,959
; EARLIER FILING DATE: 1995-06-28
; NUMBER OF SEQ ID NOS: 10
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 4
; TYPE: PRT
; ORGANISM: Escherichia coli
US-08-978-878-4

Query Match      66.9%; Score 514; DB 13; Length 151;
Best Local Similarity 68.9%; Pred. No. 6.7e-46;
Matches 104; Conservative 18; Mismatches 29; Indels 0; Gaps 0;

QY 1 MKLLKVAFAAIVVSGSALAGVVPQWGGGNGHGGNSGGPDSTLSIYQYGSANAALALQ 60
Db 1 MKLLKVAFAAIVVSGSALAGVVPQWGGGNGHGGNSGGPDSTLSIYQYGSANAALALQ 60
QY 61 SDARKSETTITSGYNGGADVGCGADNSTIELTQGFRRNATYDQLVTRVVTHEVAHAGG 120
Db 61 TDARNEDLTITQHGNGGADVGCGDDSSIDLTRQFGNSATLDWNGKNSMTVKQFGG 120
QY 121 NNAALVNOTASDSSVMVROVFGNNATANQY 151
Db 121 GNGAAYDQIASNSSVNTQVFGNNATANQY 151
```

Search completed: August 2, 2004, 15:26:47  
Job time : 168.9 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: August 2, 2004, 14:49:38 ; Search time 17.8 Seconds  
(without alignments)  
888.146 Million cell updates/sec

Title: US-09-543-407-30

Perfect score: 768

Sequence: 1 MKLLKVAFAAIVVSGSALA.....DSSVMVROVGFNNATANQY 151

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 601315 seqs, 104695340 residues

Total number of hits satisfying chosen parameters: 601315

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Pending Patents AA New.\*

- 1: /cgn2\_6/prodata/2/paa/PCT\_NEW\_COMB.pcp.\*
- 2: /cgn2\_6/prodata/2/paa/US06\_NEW\_COMB.pcp.\*
- 3: /cgn2\_6/prodata/2/paa/US07\_NEW\_COMB.pcp.\*
- 4: /cgn2\_6/prodata/2/paa/US08\_NEW\_COMB.pcp.\*
- 5: /cgn2\_6/prodata/2/paa/US09\_NEW\_COMB.pcp.\*
- 6: /cgn2\_6/prodata/2/paa/US10\_NEW\_COMB.pcp.\*
- 7: /cgn2\_6/prodata/2/paa/US60\_NEW\_COMB.pcp.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	514	66.9	151	5	US-09-741-873C-4
2	436	56.8	131	5	US-09-741-873C-2
3	101.5	13.2	520	6	US-10-479-638-21
4	93.5	12.2	256	6	US-10-425-115-301334
5	93	12.1	258	6	US-10-425-115-300390
6	92	12.0	295	6	US-10-425-115-312468
7	91.5	11.9	179	6	US-10-425-115-346132
8	91.5	11.9	299	6	US-10-170-205E-35751
9	91.5	11.9	1203	6	US-10-170-205E-741
10	91.5	11.9	1327	1	PCT-US04-02338-49
11	88.5	11.5	197	5	US-10-425-115-304391
12	88.5	11.5	388	5	US-10-425-115-304391
13	87.5	11.4	573	7	US-09-248-796A-17306
14	87.5	11.4	573	7	US-09-248-796A-17306
15	86.5	11.3	193	6	US-10-425-115-254240
16	86	11.2	131	7	US-10-425-115-311109
17	86	11.2	131	7	US-10-425-115-311109
18	85	11.1	321	1	PCT-US04-0565A-590
19	85	11.1	511	6	US-10-425-115-320950
20	85	11.1	841	7	US-09-579-062-7906
21	85	11.1	841	7	US-09-579-062-7906
22	84.5	11.0	412	7	US-09-579-062-7906
23	84.5	11.0	412	7	US-09-579-062-7906
24	84	10.9	556	6	US-10-425-115-337674
25	84	10.9	586	1	PCT-US03-24982A-317
26	83.5	10.9	586	7	US-09-579-062-7906

27	83.5	10.9	758	7	US-60-579-062-9194	Sequence 9194, Ap
28	83	10.8	376	6	US-10-491-733-2	Sequence 2, Appli
29	83	10.8	719	5	US-09-248-796A-17559	Sequence 17559, A
30	82.5	10.7	630	7	US-60-581-351-9810	Sequence 9810, Ap
31	82	10.7	279	6	US-10-425-115-343835	Sequence 343835,
32	82	10.7	443	6	US-10-100-683-7608	Sequence 7608, Ap
33	82	10.7	1358	6	US-10-778-804-11	Sequence 11, Appl
34	81.5	10.6	508	6	US-10-425-115-285216	Sequence 285216,
35	81.5	10.6	532	6	US-10-425-115-309662	Sequence 309662,
36	81	10.5	201	6	US-10-425-115-343526	Sequence 343526,
37	81	10.5	376	6	US-60-581-351-1872	Sequence 1872, Ap
38	81	10.5	1127	7	US-10-854-439-511	Sequence 511, App
39	80.5	10.5	246	6	US-10-425-115-190187	Sequence 190187,
40	80.5	10.5	269	6	US-60-581-351-9806	Sequence 9806, Ap
41	80.5	10.5	630	7	US-60-581-351-9807	Sequence 9807, Ap
42	80.5	10.5	631	7	US-10-864-138-2	Sequence 2, Appli
43	80.5	10.5	1571	6	US-10-425-115-351875	Sequence 351875,
44	80	10.4	132	6	US-10-425-115-351875	Sequence 351875,
45	80	10.4	1059	1	PCT-US03-24982A-343	Sequence 343, App

ALIGNMENTS

RESULT 1

US-09-741-873C-4  
; Sequence 4, Application US/09741873C  
; GENERAL INFORMATION:  
; APPLICANT: Olsen, Arne  
; TITLE OF INVENTION: Fibronectin Binding Protein As Well As Its Preparation  
; FILE REFERENCE: 012889-084  
; CURRENT APPLICATION NUMBER: US/09/741.873C  
; CURRENT FILING DATE: 2000-12-22  
; PRIOR APPLICATION NUMBER: SE 8801723-1  
; PRIOR FILING DATE: 1998-05-06  
; PRIOR APPLICATION NUMBER: US 08/978,878  
; PRIOR FILING DATE: 1997-11-26  
; PRIOR APPLICATION NUMBER: US 07/347,189  
; PRIOR FILING DATE: 1989-05-04  
; PRIOR APPLICATION NUMBER: US 07/789,437  
; PRIOR FILING DATE: 1991-11-06  
; PRIOR APPLICATION NUMBER: US 07/970,846  
; PRIOR FILING DATE: 1992-11-03  
; PRIOR APPLICATION NUMBER: US 08/187,865  
; PRIOR FILING DATE: 1994-01-28  
; PRIOR APPLICATION NUMBER: US 08/318,519  
; PRIOR FILING DATE: 1994-10-05  
; NUMBER OF SEQ ID NOS: 11  
; SOFTWARE: PatentIn version 3.0  
; SEQ ID NO 4  
; LENGTH: 151  
; TYPE: PRT  
; ORGANISM: Escherichia coli  
US-09-741-873C-4

Query Match	66.9%	Score 514;	DB 5;	Length 151;
Best Local Similarity	68.9%	Pred. No. 3.7e-37;		
Matches 104;	Conservative 18;	Mismatches 29;	Indels 0;	Gaps 0;
QY	1	MKLLKVAFAAIVVSGSALAIVPQWGGGNNHNGGNSGPDSTLSIYQYGSANAALQ	60	
Db	1	MKLLKVAFAAIVVSGSALAIVPQWGGGNNHNGGNSGPDSTLSIYQYGSANAALQ	60	
QY	61	SDARKSETTITQSGYNGADVGADNNTLTGTGFRNNATYDQLVTRVVVTHMAHAGG	120	
Db	61	TDARNSDLTITQHGCGGADVGQSDSSDLTGFGNSATLDWNGKSNEMTVKQFGG	120	
QY	121	NNAALVQNTASDSSVMVROVGFNNATANQY	151	
Db	121	NGRAAVQNTASNSVNTQVGFNNATANQY	151	

Query Match 13.2%: Score 101.5; DB 6; Length 520;

FEATURE:  
NAME/KEY: unsure  
LOCATION: (1)..(258)  
OTHER INFORMATION: unsure at all Xaa locations  
FEATURE:  
OTHER INFORMATION: Clone ID: MRT4577\_37025C.1.pep  
US-10-425-115-300390

Query Match 12.1%; Score 93; DB 6; Length 258;  
Best Local Similarity 26.8%; Pred. No. 1.7;  
Matches 33; Conservative 16; Mismatches 42; Indels 32; Gaps 5;

QY 17 SALAGVVPWGGGNGHNGGNSGPDSTLSIYQGSANAALQSDARKSETTITQSG-- 74  
DB 78 SSIAG-----GGGGGGGGGGTNGSGGGGGGGGSSSTAASGPGSGNVADAEGKAG 132  
QY 75 -----YNGGA--DVGGGADNSTIEL--TQNGFRNNATYDQLVTRVVTHEMAHAGG 120  
DB 133 GGMGGGANGAYSGAGGKGEGVSGVALATSSNGYNGGAAD-----ATGGG 181

QY 121 NNA 123  
DB 182 SGA 184

RESULT 6  
US-10-425-115-312468  
Sequence 312468, Application US/10425115  
GENERAL INFORMATION:  
APPLICANT: La Rosa, Thomas J.  
APPLICANT: Kovalic, David K.  
APPLICANT: Zhou, Yihua  
APPLICANT: Cao, Yongwei  
TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With  
FILE REFERENCE: 38-21(53222)B  
CURRENT APPLICATION NUMBER: US/10/425,115  
CURRENT FILING DATE: 2003-04-28  
NUMBER OF SEQ ID NOS: 369326  
SEQ ID NO 312468  
LENGTH: 295  
TYPE: PRT  
ORGANISM: Zea mays  
FEATURE:  
OTHER INFORMATION: Clone ID: MRT4577\_48027C.1.pep  
US-10-425-115-312468

Query Match 12.0%; Score 92; DB 6; Length 295;  
Best Local Similarity 30.2%; Pred. No. 2.4;  
Matches 35; Conservative 13; Mismatches 46; Indels 22; Gaps 5;

QY 26 WGGGGNHNGGNSGPDSTLSIYQGSANAALQSDARKSETTITQSGYNGADYVQGA 85  
DB 138 YGGGGYSSGGYSSG-GYANGYGVSGSGNYSNAGGGYSGS-----DYGNGAASGGYA 192  
QY 86 DNSTIELTQNGFRNNATYDQLVTRVVTHEMAHAGGN-----AALNVOTASDSS 134  
DB 193 NN-----LSSGYNNGRYN-----TIGSSDGTGNGYNSYPNPGAGNYNTGGSSSS 238

RESULT 7  
US-10-425-115-346132  
Sequence 346132, Application US/10425115  
GENERAL INFORMATION:  
APPLICANT: La Rosa, Thomas J.  
APPLICANT: Kovalic, David K.  
APPLICANT: Zhou, Yihua  
APPLICANT: Cao, Yongwei  
TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With  
FILE REFERENCE: 38-21(53222)B  
CURRENT APPLICATION NUMBER: US/10/425,115  
CURRENT FILING DATE: 2003-04-28

NUMBER OF SEQ ID NOS: 369326  
SEQ ID NO 346132  
LENGTH: 179  
TYPE: PRT  
ORGANISM: Zea mays  
FEATURE:  
OTHER INFORMATION: Clone ID: MRT4577\_78839C.1.pep  
US-10-425-115-346132

Query Match 11.9%; Score 91.5; DB 6; Length 179;  
Best Local Similarity 27.7%; Pred. No. 1.5;  
Matches 39; Conservative 10; Mismatches 55; Indels 37; Gaps 6;

QY 19 LAGVVPWGGGG-----NHNGGNSGPDSTLSIYQGSANAALQSDARKSETTI 70  
DB 64 LDGLLSLGGGGLDGLLGLTGCGGKKNKAQADSGNAQEGGNA-----QEDSGNAQ--- 116  
QY 71 TQSGYNGADYVGGGADNSTIELTQNGFRNNATYDQLVTRVVTHEMAHAGGNALVN-OT 129  
DB 117 EESGNQNAAGAGA-----ENGAANNGT-----EAAGAENAAAGNGQA 155  
QY 130 ASDSSVMVRQVGGNNATNQ 150  
DB 156 QIEGEDQAQGNAGNENAAEE 176

RESULT 8  
US-10-170-205E-35751  
Sequence 35751, Application US/10170205E  
GENERAL INFORMATION:  
APPLICANT: ADAMS, Mark  
TITLE OF INVENTION: DEVICES, SUCH AS ARRAYS, COMPRISED OF HUMAN PROTEINS OR PROTEIN  
FILE OF INVENTION: CAPTURE AGENTS, AND USES THEREOF  
FILE REFERENCE: CL001381  
CURRENT APPLICATION NUMBER: US/10/170,205E  
CURRENT FILING DATE: 2002-06-13  
NUMBER OF SEQ ID NOS: 40312  
SOFTWARE: Patent in version 3.2  
NUMBER OF SEQ ID NOS: 35751  
LENGTH: 299  
TYPE: PRT  
ORGANISM: Homo sapiens  
US-10-170-205E-35751

Query Match 11.9%; Score 91.5; DB 6; Length 299;  
Best Local Similarity 29.4%; Pred. No. 2.6;  
Matches 42; Conservative 17; Mismatches 65; Indels 19; Gaps 6;

QY 6 VAAPAAI-VTSGSALAGVVPWGGGNGHNGGNSGPDSTLSIYQGSANAALQSDAR 64  
DB 99 VAAAPVVPVAVSTSSAAGVAPNPAGSGNSNPFSSSSPTSS--SSSSPSPGSSLAESPEAA 157  
QY 65 KSETTIT-----QSGYNGADYVGGGADNSTIELTQNGFRNNATYDQLVTRVVTHEMAHAGG 120  
DB 158 GVSSTAPLPGCAAGPGTGVPAVSGALRELLACRNGD-----VSRV--KRLVDAAN 206  
QY 121 NNAALNVOTASDSSVMVRQVGF 143  
DB 207 VNAK--DVAGRKSSPLHFAAGFG 227

RESULT 9  
US-10-170-205E-741  
Sequence 741, Application US/10170205E  
GENERAL INFORMATION:  
APPLICANT: ADAMS, Mark  
TITLE OF INVENTION: DEVICES, SUCH AS ARRAYS, COMPRISED OF HUMAN PROTEINS OR PROTEIN  
FILE OF INVENTION: CAPTURE AGENTS, AND USES THEREOF  
FILE REFERENCE: CL001381  
CURRENT APPLICATION NUMBER: US/10/170,205E  
CURRENT FILING DATE: 2002-06-13  
NUMBER OF SEQ ID NOS: 40312  
SOFTWARE: Patent in version 3.2





Db 240 GAGFGDNTSSY 252

## RESULT 13

US-60-565-632-7907  
; Sequence 7907, Application US/60565632

; GENERAL INFORMATION:

; APPLICANT: Monsanto Technology, LLC

; APPLICANT: Baum, James A

; APPLICANT: Kovalic, David K.

; APPLICANT: Larosa, Thomas J.

; APPLICANT: Lu, Maolong

; APPLICANT: Munyikwa, Tichifa R. I.

; APPLICANT: Roberts, James K.

; APPLICANT: Wu, Wei

; APPLICANT: Zhang, Bei

; TITLE OF INVENTION: Methods for Genetic Control of Insect Infestations in Plants and

; FILE REFERENCE: 38-21(53403)B

; CURRENT APPLICATION NUMBER: US/60/565,632

; CURRENT FILING DATE: 2004-04-27

; NUMBER OF SEQ ID NOS: 15449

; SOFTWARE: PatentIn version 3.2

; SEQ ID NO 7907

; TYPE: PRT

; ORGANISM: Diabrotica virgifera

US-60-565-632-7907

Query Match 11.4%; Score 87.5; DB 7; Length 573;

Best Local Similarity 26.0%; Pred. No. 12;

Matches 34; Conservative 16; Mismatches 54; Indels 27; Gaps 5;

QY 30 GNHNGG--NSSGPDSTLSIYQGSANAALA-----LQSDARKSETTITQSGYNGAD 80

Db 241 GNENGTAENANADAQTQVAQ--GSTNEANAENANADVQNDAAQAANENGAAENSGNAD 299

QY 81 VGGADN-STIELTON-----GFRNNATYDQLVTRVVTHEMAHAGGNA 123

Db 300 AAOGTDTGAAAGTGNADPAQGNNGAAENSGNENGTAENANADVQNDAAQAQVNDNGA 359

QY 124 ALVNOTASDSS 134

Db 360 AAEENGNADAA 370

## RESULT 14

US-60-579-062-7907

; Sequence 7907, Application US/60579062

; GENERAL INFORMATION:

; APPLICANT: Baum, James A

; APPLICANT: Kovalic, David K

; APPLICANT: Larosa, Thomas J

; APPLICANT: Lu, Maolong

; APPLICANT: Munyikwa, Tichifa R. I.

; APPLICANT: Roberts, James K

; APPLICANT: Wu, Wei

; APPLICANT: Zhang, Bei

; TITLE OF INVENTION: Methods for Genetic Control of Insect Infestations in Plants and

; FILE REFERENCE: 38-21(53403) C

; CURRENT APPLICATION NUMBER: US/60/579,062

; CURRENT FILING DATE: 2004-06-11

; NUMBER OF SEQ ID NOS: 41445

; SOFTWARE: PatentIn version 3.2

; SEQ ID NO 7907

; TYPE: PRT

; ORGANISM: Diabrotica virgifera

US-60-579-062-7907

Query Match

Best Local Similarity 11.4%; Score 87.5; DB 7; Length 573;

Matches 34; Conservative 16; Mismatches 54; Indels 27; Gaps 5;

Matches 34; Conservative 16; Mismatches 54; Indels 27; Gaps 5;

QY 30 GNHNGG--NSSGPDSTLSIYQGSANAALA-----LQSDARKSETTITQSGYNGAD 80

Db 241 GNENGTAENANADAQTQVAQ--GSTNEANAENANADVQNDAAQAANENGAAENSGNAD 299

QY 81 VGGADN-STIELTON-----GFRNNATYDQLVTRVVTHEMAHAGGNA 123

Db 300 AAOGTDTGAAAGTGNADPAQGNNGAAENSGNENGTAENANADVQNDAAQAQVNDNGA 359

QY 124 ALVNOTASDSS 134

Db 360 AAEENGNADAA 370

## RESULT 15

US-10-425-115-254240

; Sequence 254240, Application US/10425115

; GENERAL INFORMATION:

; APPLICANT: La Rosa, Thomas J.

; APPLICANT: Kovalic, David K.

; APPLICANT: Zhou, Yihua

; APPLICANT: Cao, Yongwei

; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With

; FILE REFERENCE: 38-21(53222)B

; CURRENT APPLICATION NUMBER: US/10/425,115

; CURRENT FILING DATE: 2003-04-28

; NUMBER OF SEQ ID NOS: 369326

; SEQ ID NO 254240

; LENGTH: 193

; TYPE: PRT

; ORGANISM: Zea mays

; FEATURE:

; OTHER INFORMATION: Clone ID: MRT4577\_163446C.1.pep

US-10-425-115-254240

Query Match 11.3%; Score 86.5; DB 6; Length 193;

Best Local Similarity 26.9%; Pred. No. 4.3;

Matches 35; Conservative 15; Mismatches 53; Indels 27; Gaps 5;

QY 27 GGGGNHNGG---GNSSGPDSTLSIYQGSANAALALQSDARKSETTITQSGYNGADVQ 83

Db 76 GGGASQNGGGYSGSGSGSGSTYSQGGYSGYSSNA-----GGTGGGGGGGQ 126

QY 84 --GADNSTIELTQNGFRNNATYDQLVTRVVTHEMAHAGGNAALVNQT-----ASDSS 134

Db 127 AGGYWNSNAQSGSGTSGSYANRYWDGSEGGANANGCGGNGSQNGGGGGSGDGS 186

QY 135 VMVRQVGFN 144

Db 187 -----GYGN 190

Search completed: August 2, 2004, 15:29:56

Job time : 18.8 secs

This Page Blank (uspto)

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: August 2, 2004, 14:39:53 ; Search time 9.4 Seconds  
(without alignments)  
1545.204 Million cell updates/sec

Title: US-09-543-407-30  
Perfect score: 768  
Sequence: 1 MKLLKVAFAAIVVSGSALA.....DSSVWVRQVGFNNATANQY 151

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 283366 seqs, 96191526 residues

Total number of hits satisfying chosen parameters: 283366

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : PIR 78:\*  
1: PIR1:\*  
2: PIR2:\*  
3: PIR3:\*  
4: PIR4:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	681	88.7	151	JC6039	fimbrin protein ag
2	681	88.7	151	A10635	major curlin chain
3	517	67.3	151	S70788	curlin protein csg
4	502.5	65.4	152	D90806	curlin major subun
5	502.5	65.4	152	H85665	hypothetical prote
6	104.5	13.6	151	S70787	curlin nucleator p
7	104.5	13.6	151	C90806	minor curlin subun
8	104.5	13.6	151	G85665	curlin minor chain
9	101	13.2	1748	S42136	cnjB protein - Tet
10	99.5	13.0	151	JC6040	fimbrin protein ag
11	99.5	13.0	151	AH0635	nucleation compone
12	99.5	13.0	2174	E95965	hypothetical glyci
13	98	12.8	145	AD3143	conserved hypotet
14	98	12.8	145	H98144	hypothetical prote
15	97.5	12.7	645	S70825	probable PPE prote
16	96.5	12.6	313	T04776	hypothetical prote
17	94	12.2	354	B70663	probable PPE prote
18	94	12.2	656	D96831	hypothetical prote
19	93.5	12.2	256	T03371	glycine-rich prote
20	93.5	12.2	2204	A70524	probable PPE prote
21	92.5	12.0	1028	A56038	DNA-binding protei
22	92.5	12.0	1213	S16356	ovo protein - frui
23	92	12.0	575	S35327	protein kinase S99
24	91.5	11.9	409	T20847	hypothetical prote
25	91.5	11.9	573	C86266	F3F19.21 protein -
26	91.5	11.9	1910	AF0394	probable adhesin h
27	91	11.8	652	E97857	cell surface antig
28	90.5	11.8	447	G84687	probable disease r
29	89.5	11.7	963	B70524	probable PPE prote

30	89.5	11.7	1238	2	AH0038	probable exported
31	89	11.6	343	2	T05221	hypothetical prote
32	89	11.6	959	2	B44402	nuclear pore compl
33	88.5	11.5	967	2	S66852	hypothetical prote
34	88.5	11.5	1217	2	S52714	sericin1B - silkwo
35	88	11.5	407	2	T21956	hypothetical prote
36	88	11.5	586	2	T26667	hypothetical prote
37	88	11.5	1778	2	T50074	probable nucleopor
38	87.5	11.4	1053	2	B70987	probable PPE prote
39	87	11.3	590	2	E70946	probable PPE prote
40	87	11.3	3624	2	AD0835	large repetitive p
41	86.5	11.3	262	2	S00275	tail fiber protein
42	86.5	11.3	590	1	A45621	leishmanolysin (EC
43	86.5	11.3	745	2	E64559	outer membrane pro
44	86	11.2	582	2	F70675	probable PPE prote
45	86	11.2	639	2	C42049	leishmanolysin (EC

ALIGNMENTS

RESULT 1

JC6039  
fimbrin protein agfa precursor - Salmonella enteritidis  
C:Species: Salmonella enteritidis  
C:Date: 31-Dec-1996 #sequence revision 31-Dec-1996 #text\_change 08-Oct-1999  
C:Accession: JC6039; PC6015; A44898  
R:Collinson, S.K.; Clouthier, S.C.; Doran, J.L.; Bansen, P.A.; Kay, W.W.  
J. Bacteriol. 178, 662-667, 1996  
A:Title: Salmonella enteritidis agfBAC operon encoding thin, aggregative fimbrae.  
A:Reference number: JC6039; MUID:96146512; PMID:8550497  
A:Accession: JC6039  
A:Molecule type: DNA  
A:Residues: 1-151 <COL>  
A:Cross-references: GB:U43280; NID:g1184712; PIDN:AAC43599.1; PID:g1184714  
A:Accession: PC6015  
A:Molecule type: protein  
A:Residues: 21-52 <CO2>  
A:Experimental source: strain 27655-3b  
A:Note: the authors translated the codon ACG for residue 44 as Ile  
R:Collinson, S.K.; Emody, L.; Muller, K.H.; Trust, T.J.; Kay, W.W.  
J. Bacteriol. 173, 4773-4781, 1991  
A:Title: Purification and characterization of thin, aggregative fimbrae from Salmonella  
A:Reference number: A44898; MUID:91310586; PMID:1677357  
A:Contents: 27655  
A:Accession: A44898  
A>Status: preliminary  
A:Molecule type: protein  
A:Residues: 21-33 <CO3>  
A:Note: sequence extracted from NCBI backbone (NCBIP:45936)  
C:Genetics:  
A:Gene: agfa  
C:Function:  
A:Description: major component of thin aggregative fimbrae  
A:Note: fimbrae bind to fibronectin, plasminogen, tissue plasminogen activator  
C:Keywords: fimbria  
F:1-20/Domain: signal sequence #status predicted <SIG>  
F:21-151/Product: fimbrin protein agfa #status experimental <MAT>

Query Match 88.7%; Score 681; DB 2; Length 151;  
Best Local Similarity 90.7%; Pred. No. 3.3e-49;  
Matches 137; Conservative 2; Mismatches 12; Indels 0; Gaps 0;

QY	1	MKLLKVAFAAIVVSGSALA	GVFQWGGGNGHNGGNSG	SPDSTLSIYQGSANAALAQ	60
DB	1	MKLLKVAFAAIVVSGSALA	GVFQWGGGNGHNGGNSG	SPDSTLSIYQGSANAALAQ	60
QY	61	SDARSETTITQSGVNGADY	VGQADNSTIELTQGNPNNA	TYDQLVTRVVVTHMAHAGG	120
DB	61	SDARSETTITQSGVNGADY	VGQADNSTIELTQGNPNNA	TYDQLVTRVVVTHMAHAGG	120
QY	121	NNAALVNQTS	DSSVWVRQVGFNNATANQY	151	



A;Experimental source: strain K-12, substrain MG1655  
C;Genetics:  
A;Gene: csgB  
A;Map position: 23.15  
C;Function:  
A;Description: minor component of wild-type curli; interaction between CsgA and CsgB tri-  
A;Note: curli are thin, coiled fibers expressed on the surface of *Escherichia coli* that i  
and H-kinogen; in the absence of CsgA, CsgB can self-assemble into polymers  
F;1-21/Domain: signal sequence #status predicted <SIG>  
F;2-151/Product: minor curlin chain #status predicted <MAT>

Query Match 13.6%; Score 104.5; DB 2; Length 151;  
Best Local Similarity 29.1%; Pred. No. 0.087;  
Matches 34; Conservative 15; Mismatches 57; Indels 11; Gaps 3;

Qy 38 SSGPDSLTSLIYQGSANAALQSDARKSETTITOSGYNGADVCGGADNSTIELTQNGF 97  
Db 21 AAGDYDLANSEYNF----AVNELSKSFNQAAIIQAGTNNSQLRQGGSKLLAVVAQEGS 76  
Qy 98 RNNATYDQLVTRVVTTHMAH---AGGNNAAALVNQTASDSSVMVRQVFGNNATANQY 151  
Db 77 SNRAKIDQ---TGDYNLAVIDQAGSANDASISQGYGNTAMIIQKSGGNKANTQY 129

RESULT 7  
C90806  
minor curlin subunit precursor CsgB [imported] - *Escherichia coli* (strain O157:H7, subst  
C;Species: *Escherichia coli*  
C;Date: 18-Jul-2001 #sequence\_revision 18-Jul-2001 #text\_change 18-Jul-2001  
C;Accession: C90806  
R;Hayashi, T.; Makino, K.; Ohnishi, M.; Kurokawa, K.; Ishii, K.; Yokoyama, K.; Han, C.G.  
gasawara, N.; Yasunaga, T.; Kuhara, S.; Shiba, T.; Hattori, M.; Shinagawa, H.  
DNA Res. 8, 11-22, 2001  
A;Title: Complete genome sequence of enterohemorrhagic *Escherichia coli* O157:H7 and geno  
A;Reference number: A99629; UID:21156231; PMID:11258796  
A;Accession: C90806  
A;Status: preliminary  
A;Molecule type: DNA  
A;Residues: 1-151 <HAY>  
A;Cross-references: GB:BA000007; PIDN:BAB34842.1; PID:g13360879; GSPDB:GN00154  
A;Experimental source: strain O157:H7, substrain RMD 0509952  
C;Genetics:  
A;Gene: Ecs1419

Query Match 13.6%; Score 104.5; DB 2; Length 151;  
Best Local Similarity 29.1%; Pred. No. 0.087;  
Matches 34; Conservative 15; Mismatches 57; Indels 11; Gaps 3;

Qy 38 SSGPDSLTSLIYQGSANAALQSDARKSETTITOSGYNGADVCGGADNSTIELTQNGF 97  
Db 21 AAGDYDLANSEYNF----AVNELSKSFNQAAIIQAGTNNSQLRQGGSKLLAVVAQEGS 76  
Qy 98 RNNATYDQLVTRVVTTHMAH---AGGNNAAALVNQTASDSSVMVRQVFGNNATANQY 151  
Db 77 SNRAKIDQ---TGDYNLAVIDQAGSANDASISQGYGNTAMIIQKSGGNKANTQY 129

RESULT 8  
G85665  
curlin minor chain precursor, CsgA homolog [imported] - *Escherichia coli* (strain O157:H7  
C;Species: *Escherichia coli*  
C;Date: 16-Feb-2001 #sequence\_revision 16-Feb-2001 #text\_change 14-Sep-2001  
C;Accession: G85665  
R;Perna, N.T.; Plunkett III, G.; Burland, V.; Mau, B.; Glasner, J.D.; Rose, D.J.; Mayhew  
Miller, L.; Grotbeck, E.G.; Davis, N.W.; Lim, A.; Dimalanta, E.; Potamousis, K.; Apodaca,  
Nature 409, 529-533, 2001  
A;Title: Genome sequence of enterohemorrhagic *Escherichia coli* O157:H7.  
A;Reference number: A85480; UID:21074935; PMID:11206551  
A;Accession: G85665  
A;Status: preliminary  
A;Molecule type: DNA  
A;Residues: 1-151 <STO>  
A;Cross-references: GB:AE005174; NID:g12514573; PIDN:AAG55787.1; GSPDB:GN00145; UWGP:212

A;Experimental source: strain O157:H7, substrain EDL933

C;Genetics:

A;Gene: cs9B

Query Match 13.6%; Score 104.5; DB 2; Length 151;

Best Local Similarity 29.1%; Pred. No. 0.087;

Matches 34; Conservative 15; Mismatches 57; Indels 11; Gaps 3;

QY 38 SSGPDTLSIYQGSNAALALOSDARKSETTITQSGYNGADVGCGADNSFIETQNGF 97

DB 21 AAGDYLANSEYNP-----AVNELSKSFNQAAITGQAGTNNASQLRGGSKLLAVVAQGS 76

QY 98 RNNATVDQLVTRVVTTHMAH---AGGNAALVNQTASDSSVMVRQVGFNNATANQY 151

DB 77 SNRAKIDQ----TGDYNLAYIDQAGSANDASISQAGYNTAMIIQKSGSKNANITQY 129

RESULT 9

S42136

CnJB protein - Tetrahymena thermophila

C;Species: Tetrahymena thermophila

C;Date: 19-Mar-1997 #sequence\_revision 18-Jul-1997 #text\_change 07-Dec-1999

C;Accession: S42136; S42135; S03650

R;Taylor, F.M.; Martindale, D.W.

submitted to the EMBL Data Library, October 1992

A;Reference number: S42136

A;Accession: S42136

A;Molecule type: DNA

A;Residues: 1-1748 <TAY>

A;Cross-references: EMBL:L03710; NID:g161751; PID:g161752

R;Taylor, F.M.; Martindale, D.W.

Nucleic Acids Res. 21, 4610-4614, 1993

A;Title: Retroviral-type zinc fingers and glycine-rich repeats in a protein encoded by c

A;Reference number: S42135; MUID:94051569; PMID:8233798

A;Accession: S42135

A;Molecule type: DNA

A;Residues: 1164-1174; 1179-1198; 1233-1252; 1285-1293; 1297-1309; 1316-1326; 1331-1341; 1343-1

A;Cross-references: EMBL:L03710

R;Martindale, D.W.; Taylor, F.M.

Nucleic Acids Res. 16, 2189-2201, 1988

A;Title: Multiple introns in a conjugation-specific gene from Tetrahymena thermophila.

A;Reference number: S03650; MUID:88189811; PMID:3357771

A;Accession: S03650

A;Molecule type: DNA

A;Residues: 236-250, '1', 252-255, 'N', 257-773 <MAR>

A;Cross-references: EMBL:X06462

C;Genetics:

A;Gene: cnJB

A;Genetic code: SGC5

A;Introns: 85/3; 136/1; 157/3; 201/2; 250/2; 327/3; 499/1; 573/2; 607/3; 708/3; 777/3; 8

C;Keywords: zinc finger

F;1164-1450/Region: glycine-rich

F;1451-1464/Region: zinc finger CCHC motif

F;1478-1491/Region: zinc finger CCHC motif

F;1501-1514/Region: zinc finger CCHC motif

F;1530-1543/Region: zinc finger CCHC motif

F;1555-1568/Region: zinc finger CCHC motif

F;1579-1592/Region: zinc finger CCHC motif

F;1602-1615/Region: zinc finger CCHC motif

F;1626-1748/Region: glycine-rich

Query Match 13.2%; Score 101.; DB 2; Length 1748;

Best Local Similarity 29.7%; Pred. No. 2.6;

Matches 38; Conservative 18; Mismatches 32; Indels 40; Gaps 7;

QY 25 QWGGGHNHGG---GNSSPDTLSIYQGSNAALALOSDARKSETTITQSGYNGADV 81

DB 1640 QFGGGNSGGSGWGTSSGSDWN-----CQSNVQES-TTTSQGGWS--- 1680

QY 82 GQAGNSTT-ELTQGFNNATVDQLVTRVVTTHMAHAGGNAALVNQTASDSSVMVRQVG 141

DB 1681 -SGSGNQ----TGGGNSNDNQOQ-----QNTGGGWSGSSNQNTWSS----- 1722

QY 142 FGNNTAN 149

DB 1723 WGSNNQAS 1730

RESULT 10

JC6040

fimbrin protein agfB precursor - Salmonella enteritidis

C;Species: Salmonella enteritidis

C;Date: 31-Dec-1996 #sequence\_revision 31-Dec-1996 #text\_change 08-Oct-1999

C;Accession: JC6040

R;Collinson, S.K.; Clouthier, S.C.; Doran, J.L.; Baner, P.A.; Kay, W.W.

J. Bacteriol. 178, 662-667, 1996

A;Title: Salmonella enteritidis agfBAC operon encoding thin, aggregative fimbriae.

A;Reference number: JC6039; MUID:96146512; PMID:8550497

A;Accession: JC6040

A;Molecule type: DNA

A;Residues: 1-151 <COL>

A;Cross-references: GB:U43280; NID:g1184712; PID:AAC43598.1; PID:g1184713

A;Experimental source: strain 276755-3b

C;Genetics:

A;Gene: agfB

C;Function:

A;Description: minor component of thin aggregative fimbriae

A;Note: fimbriae bind to fibronectin, plasminogen, tissue plasminogen activator

C;Keywords: fimbria

F;1-21/Domain: signal sequence #status predicted <SIG>

F;22-151/Product: fimbrin protein agfB #status predicted <MAT>

Query Match 13.0%; Score 99.5; DB 2; Length 151;

Best Local Similarity 28.2%; Pred. No. 0.22; 51; Indels 17; Gaps 4;

Matches 33; Conservative 16; Mismatches 17; Gaps 4;

QY 51 GSANAALALOSDARKSE-----TTITQSGYNGADVQ--GADNST-----IELTQ 94

DB 14 GAPGIATATNYDLARSEYNFANVELSKSFNQAAIIQVGTDSARVRQEGSKLLSVISQ 73

QY 95 NGFNATYQLVTRVVTTHMAHAGGNAALVNQTASDSSVMVRQVGFNNATANQY 151

DB 74 EGGNNRAKVDQAGNYNPFAY-IEQTGNANDASISQAYGNSAAIIQKSGGNKANITQY 129

RESULT 11

AH0635

nucleation component of curlin monomers [imported] - Salmonella enterica subsp. enterica

C;Species: Salmonella enterica subsp. enterica serovar typhi

A;Note: this species has also been called Salmonella typhi

C;Date: 09-Nov-2001 #sequence\_revision 09-Nov-2001 #text\_change 18-Nov-2002

C;Accession: AH0635

R;Parkhill, J.; Dougan, G.; James, K.D.; Thomson, N.R.; Pickard, D.; Wain, J.; Churcher,

th, T.; Connerton, P.; Crozin, A.; Davis, P.; Davies, R.M.; Dowd, L.; White, N.; Farrar,

S.; Moule, S.; O'Gaora, P.

Nature 413, 848-852, 2001

A;Authors: Parry, C.; Quail, M.; Rutherford, K.; Simmonds, M.; Skelton, J.; Stevens, K.;

A;Title: Complete genome sequence of a multiple drug resistant Salmonella enterica serov

A;Reference number: AB0502; MUID:21534947; PMID:11677508

A;Accession: AH0635

A;Status: preliminary

A;Molecule type: DNA

A;Residues: 1-151 <PAR>

A;Cross-references: GB:AL513382; PIDN:CAD08267.1; PID:g16502314; GSPDB:GN00176

C;Genetics:

A;Gene: STV1180

Query Match

Best Local Similarity 13.0%; Score 99.5; DB 2; Length 151;

Matches 33; Conservative 16; Mismatches 17; Indels 17; Gaps 4;

QY 51 GSANAALALOSDARKSE-----TTITQSGYNGADVQ--GADNST-----IELTQ 94

DB 14 GAPGIATATNYDLARSEYNFANVELSKSFNQAAIIQVGTDSARVRQEGSKLLSVISQ 73

QY 95 NGFNATYQLVTRVVTTHMAHAGGNAALVNQTASDSSVMVRQVGFNNATANQY 151

DB 74 EGENRAKVQAGNYFAY-IEOTGHANDASISQSYAGNSAIIQSGNKANITQY 129

RESULT 12

E55965  
hypothetical glycine-rich protein [imported] - Sinorhizobium meliloti (strain 1021) magA  
C:Species: Sinorhizobium meliloti  
C:Date: 24-Aug-2001 #sequence\_revision 24-Aug-2001 #text\_change 30-Sep-2001  
C:Accession: E55965  
R:Finan, T.M.; Weidner, S.; Wong, K.; Buhrmester, J.; Chain, P.; Vorholter, F.J.; Hernan  
Proc. Natl. Acad. Sci. U.S.A. 98, 9889-9894, 2001  
A:Title: The complete sequence of the 1.683-kb pSYMB megaplasmid from the N2-fixing endo  
A:Reference number: A95842; MUID:21396508; PMID:11481431  
A:Accession: E55965  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-2174 <XUP>  
A:Cross-references: GB:AL591985; PIDN:CAC49389.1; PID:gi51540875; GSPDB:GNO0167  
R:Experimental source: strain 1021, megaplasmid pSYMB  
R:Galibert, F.; Finan, T.M.; Long, S.R.; Puhler, A.; Abola, P.; Ampe, F.; Barloy-Hubler,  
peta, D.; Chain, P.; Cowie, A.; Davis, R.W.; Dreano, S.; Federspiel, N.A.; Fisher, R.F.;  
L.; Hyman, R.W.; Jones, T.  
Science 293, 668-672, 2001  
A:Authors: Kahn, D.; Kahn, M.L.; Kalman, S.; Keating, D.H.; Kiss, B.; Komp, C.; Lelaure,  
heault, P.; Vandenbol, M.; Vorholter, F.J.; Weidner, S.; Wells, D.H.; Wong, K.; Yeh, K.;  
A:Title: The composite genome of the legume symbiont Sinorhizobium meliloti.  
A:Reference number: A96039; MUID:21368234; PMID:11474104  
C:Genetics: annotation  
C:Comments:  
A:Gene: SMD21548  
A:Genome: plasmid

```
Query Match      13.0%; Score 99.5; DB 2; Length 2174;
Best Local Similarity 23.0%; Pred. No. 4.3;
Matches 32; Conservative 14; Mismatches 66; Indels 27; Gaps 2;
```

QY	11	AIVSGSALAGVVFQWGGGNGHNGSSGPDSTLIYQVGSNAPALALQSDPKSETTI	70
	:	:::::	:
Db	892	AVFTTKGDSHGLLAQSIGGGGNGGSASS-----MSRTAGTVDTIS	932
	:	:::::	:
QY	71	TQSGYGNGADVGQQADNSTIELTQNGFRNNATVDQLVTRVTWTMAHAGGNAAALNQTA	130
	:	:::::	:
Db	933	ADIGVSVCSSAGGAGGVATAALKDASVTTADDIYANAIVQSIGGGGNGGV-----	986
	:	:::::	:
QY	131	SDSSVMVRQVFGNNATAN	149
	:	:::::	:
Db	987	--GSVNSKEIGSGENLTAN	1003

RESULT 13  
AD3143  
conserved hypothetical protein Atu4768 [imported] - Agrobacterium tumefaciens (strain C58)  
C/Species: Agrobacterium tumefaciens  
C/Date: 11-Jan-2002 #sequence\_revision 11-Jan-2002 #text\_change 18-Nov-2002  
C/Accession: AD3143  
R:Wood, D.W.; Setubal, J.C.; Kaul, R.; Monks, D.; Chen, L.; Wood, G.E.; Chen, Y.; Woo, I.;  
erage, G.; Gillet, W.; Grant, C.; Guenther, D.; Kutayavin, T.; Levy, R.; Li, M.; McClellan,  
Karp, P.; Romero, P.; Zhang, S.  
Science 294, 2311-2323, 2001  
A/Authors: Yoo, H.; Tao, Y.; Biddle, P.; Jung, M.; Krespan, W.; Perry, M.; Gordon-Kamm,  
ster, E.W.  
A/Title: The Genome of the Natural Genetic Engineer Agrobacterium tumefaciens C58.  
A/Reference number: AB2577; MUID:21608550; PMID:11743193  
A/Accession: AD3143  
A/Status: preliminary  
A/Molecule type: DNA  
A/Residues: 1-145 <CUR>  
A/Cross-references: GB:AE068699; PIDN:AL4562.1; PID:G17743277; CSPDB:GN00187  
A/Experimental source: strain C58 (Dupont)  
C/Genetics:  
A/Gene: Atu4768  
A/Map position: linear chromosome

Query Match	12.8%;	Score 98;	DB 2;	Length 145;
Best Local Similarity	27.1%;	Pred. No. 0.29;		
Matches 38;	Conservative 25;	Mismatches 57;	Indels 20;	Gaps 5;
QY	3	LLKVAFAAIWVGSLAGVVP-----QWG-----CGGNHNGGN-----SSGPDST	44	
Db	1	MIRKFTASALVALVGLSAAAPAMANDVRIEQVHSNSAGGAQEGYGNRIETQYGGYNR	60	
QY	45	LSIVQYGSANAALQSDARKSTTTITQSGYNGADVQGGADNSTIELTQGFNNATYD	104	
Db	61	IVGHQYGRHNL-SAVGQEGHDNYGSTTQNGNRNVAGIGQFGSNHTTILTDGNGNIAAGV	119	
QY	105	QLVTRVVTTHMAHAGGNAA	124	
Db	120	Q-VGRGCSANVSQGDNDVA	138	

RESULT 14  
H98144  
hypothetical protein AGR\_L\_228 [imported] - Agrobacterium tumefaciens (strain C)  
S:Species: Agrobacterium tumefaciens  
C:Date: 22-Oct-2001 #sequence\_revision 22-Oct-2001 #text\_change 18-Nov-2002  
C:Accession: H98144  
R:Goodner, B.; Hinkle, G.; Gattung, S.; Miller, N.; Blanchard, M.; Quorllo, B.;  
A.: Liu, F.; Wolham, C.; Allinger, M.; Doughty, D.; Scott, C.; Lappas, C.; Mar-  
Science 294, 2323-2328, 2001  
A:Title: Genome Sequence of the Plant Pathogen and Biotechnology Agent Agrobact-  
A:Reference number: A97359; MUID:21608551; PMID:11743194  
A:Accession: H98144  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-445 <KUR>  
A:Cross-references: GB:AF007870; PIDN:AAK88682.1; PID:gl5158413; GSPDB:GN00170  
C:Genetics:  
A:Gene: AGR\_L\_228  
A:Map position: linear chromosome

Query Match		12.8%; Score 98; DB 2; Length 145;
Best Local Similarity	27.1%; Pred. No. 0.29;	
Matches	38; Conservative 25; Mismatches 57; Indels 20; Gaps 5;	
QY	3 LLKVNAPAAIVSSSALAGVTP-----QWG----GGNHNGGN-----SSGPDST 44 ::: : : : : : : : : : : : : : : : : :	
Dd	1 MIRKSFTASALVALVGLSAAPPANADVRIEQWNSAGGAEGYGRIITYQNGGYNR 60   : : : : : : : : : : : : : : : :	
QY	45 LSIYQYGSANAALALOSDARKSETTTITQSGYNGADVGOGADNSTIELTQTGFNNATYD 104 ::: : : : : : : : : : : : : : : : :	
Dd	61 IVGQYQYGHNLUS-AVGGECHDYGSTQTONGNENAVIGQFGSNHTTILTDGNGNIAGV 119   : : : : : : : : : : : : : : : :	
QY	105 QLTVRVVTTHMAHAGGNAA 124   : : : : : : : : : : : : : : : :	
Dd	120 Q-VGRGCSANVSOGENDNVA 138   : : : : : : : : : : : : : : : :	

RESULT 15  
F70825 probable PPE protein - Mycobacterium tuberculosis (strain H37RV)  
C:Species: Mycobacterium tuberculosis  
C:Date: 17-Jul-1998 #sequence\_revision 17-Jul-1998 #text\_change 15-Sep-2003  
C:Accession: F70825  
R:Coile, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; O'Connor, R.; Davies, R.; Devlin, K.; Feltwell, T.; Gentles, S.; Hamlin, N.; Hojajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S.; Nature 393, 537-544 1998  
A:Authors: Squares, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.  
A:Title: Deciphering the Biology of Mycobacterium tuberculosis from the complete genome sequence  
A:Reference number: A70500; MUID:98295987; PMID:9634230  
A:Accession: F70825  
A:Status: preliminary; nucleic acid sequence not shown; translation not shown  
A:Molecule type: DNA  
A:Residues: 1-645 <COL>  
A:Cross-references: GB:AL021958; GB:AL123456; NID:q3261536; PIDN:CAA17522.1; PIR:Q00000

A;Residues: 1-645 <COL>  
A;Cross-references: GB:AL021958; GB:AL123456; NID:q3261536; PIDN:CAA17522.1; PID:e1253292





GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: August 2, 2004, 14:36:12 ; Search time 5.3 Seconds  
(without alignments)  
1483.508 Million cell updates/sec

Title: US-09-543-407-30  
Perfect score: 768  
Sequence: 1 MKLLKVAFAAIVVSGSALA.....DSSVMVRQVGFNNATANQY 151

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 141681 seqs, 52070155 residues  
Total number of hits satisfying chosen parameters: 141681

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : SwissProt\_42.\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	681	88.7	151	1 CSGA_SALTY	P55225 salmonella
2	517	67.3	151	1 CSGA_ECOLI	P28307 escherichia
3	502.5	65.4	152	1 CSGA_ECO57	Q93024 escherichia
4	104.5	13.6	151	1 CSGB_ECOLI	P39828 escherichia
5	99.5	13.0	151	1 CSGB_SALTY	Q827m3 salmonella
6	92.5	12.0	1028	1 CSGB_SALTY	P55226 salmonella
7	91.5	11.9	1327	1 OVO_DROME	P15121 drosophila
8	90	11.7	1656	1 OMPB_RICUA	Q95271 homo sapien
9	89	11.6	959	1 N100_YEAST	O06853 r outer mem
10	88	11.5	1778	1 N189_SCHPO	Q02629 saccharomyc
11	86.5	11.3	262	1 N189_SCHPO	Q9utk4 schizosacch
12	86.5	11.3	590	1 GP63_LEIDO	P07875 bacterioph
13	86.5	11.2	646	1 GP63_LEIME	P23223 leishmania
14	85.5	11.1	1567	1 ICEN_XANTC	P43150 leishmania
15	85	11.1	599	1 GP63_LEICH	P15706 leishmania
16	85	11.1	602	1 GP63_LEIMA	P08148 leishmania
17	85	11.1	850	1 PRSN_PIERA	Q9ubq4 pleris rapa
18	85	11.1	655	1 OMPB_RICCN	Q9kka3 r outer mem
19	84.5	11.0	678	1 YF48_MYCTU	Q10778 mycobacteri
20	84.5	11.0	760	1 YBIL_ECOLI	Q33479 escherichia
21	84.5	11.0	1196	1 ICEN_PSESX	O34792 pseudomonas
22	84.5	11.0	1469	1 BCL9_DROME	Q96143 drosophila
23	84.5	11.0	363	1 OMPC_KLEPN	Q48473 klebsiella
24	84	10.9	576	1 DEAF_DROME	Q24180 drosophila
25	84	10.9	576	1 ICEN_PSESX	O30611 pseudomonas
26	83.5	10.9	1148	1 ICEN_PSESX	P34291 caenorhabdi
27	83	10.8	401	1 YK03_CABEL	P04480 citrobacter
28	82.5	10.7	592	1 CEA_GITER	P07184 drosophila
29	81.5	10.6	172	1 CH18_DROME	Q03648 plasmodium
30	81.5	10.6	347	1 MSA2_PLAF2	Q05925 homo sapien
31	81.5	10.6	392	1 HNE1_HUMAN	P25074 oryza sativ
32	81	10.5	165	1 GRP1_ORYSA	Q9ul36 homo sapien
33	81	10.5	1845	1 Z236_HUMAN	

34	80.5	10.5	534	1	GCR2_YEAST	Q01722 saccharomyc
35	80.5	10.5	874	1	ALB6_AZOVI	Q9zfh0 azotobacter
36	80.5	10.5	880	1	GUNA_THEFU	P26221 thermomonos
37	80.5	10.5	1210	1	ICEN_PSEFL	P09815 pseudomonas
38	80.5	10.5	1258	1	ICEN_ERWHE	P16329 erwinia her
39	80	10.4	380	1	FLYA_COLGL	Q00374 colletotric
40	80	10.4	491	1	YK98_MYCTU	Q10707 mycobacteri
41	80	10.4	850	1	PRSN_PIEBR	Q9gw36 pieris bras
42	80	10.4	1571	1	C3G_DROME	O77086 drosophila
43	80	10.4	5560	1	SPEN_DROME	O8sx83 drosophila
44	79.5	10.4	163	1	HCY_NATPH	P39442 natronomona
45	79.5	10.4	663	1	DUS8_MOUSE	O09112 mus musculu

ALIGNMENTS

RESULT 1	ID	CSGA_SALTY	STANDARD;	PRT;	151 AA.
AC	P55225;				
DT	01-OCT-1996 (Rel. 34, Created)				
DT	01-OCT-1996 (Rel. 34, Last sequence update)				
DT	01-OCT-2003 (Rel. 42, Last annotation update)				
DE	Major curlin subunit precursor (Fimbrin SEF17).				
GN	CSGA OR AGFA OR STM1144 OR STY1181 OR TL776.				
OS	Salmonella typhimurium,				
OS	Salmonella typhi, and				
OS	Salmonella enteritidis.				
OC	Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;				
OC	Enterobacteriaceae; Salmonella.				
OX	NCBI_TaxID=602, 601, 592;				
RN	[1]				
RP	SEQUENCE FROM N.A.				
RC	SPECIES=S.typhimurium; STRAIN=SR-11;				
RX	MEDLINE=98117058; PubMed=9457880;				
RA	Romling U., Bian Z., Hammar M., Sierralta W.D., Normark S.;				
RT	"Curli fibers are highly conserved between Salmonella typhimurium and				
RT	Escherichia coli with respect to operon structure and regulation.;"				
RL	J. Bacteriol. 180:722-731(1998).				
RN	[2]				
RP	SEQUENCE FROM N.A.				
RC	SPECIES=S.typhimurium; STRAIN=LT2 / SGSC1412 / ATCC 700720;				
RX	MEDLINE=21534948; PubMed=11677609;				
RA	McClelland M., Sanderson K.E., Spieth J., Clifton S.W., Latreille P.,				
RA	Courtney L., Porwollik S., Ali J., Dante M., Du F., Hou S., Layman D.,				
RA	Leonard S., Nguyen C., Scott K., Holmes A., Grewal N., Mulvaney E.,				
RA	Ryan E., Sun H., Florea L., Miller W., Stoneking T., Nhan M.,				
RA	Waterston R., Wilson R.K.;				
RT	"Complete genome sequence of Salmonella enterica serovar Typhimurium				
RT	LT2.;"				
RL	Nature 413:852-856 (2001).				
RN	[3]				
RP	SEQUENCE FROM N.A.				
RC	SPECIES=S.typhi; STRAIN=CT18;				
RX	MEDLINE=21534947; PubMed=11677608;				
RA	Parkhill J., Dougan G., James K.D., Thomson N.R., Pickard D., Wain J.,				
RA	Churcher C., Mungall K.L., Bentley S.D., Holden M.T.G., Sebahia M.,				
RA	Baker S., Basham D., Brooks K., Chillingworth T., Connor P.,				
RA	Cronin A., Davis P., Davies R.M., Dowd L., White N., Farrar J.,				
RA	Feltwell T., Hamlin N., Haque A., Hien T.T., Holroyd S., Jagers K.,				
RA	Krogh A., Larsen T.S., Leather S., Moule S., O'Gaora P., Parry C.,				
RA	Quail M.A., Rutherford K., Simmonds M., Skelton J., Stevens K.,				
RA	Whitehead S., Barrall B.G.;				
RT	"Complete genome sequence of a multiple drug resistant Salmonella				
RT	enterica serovar Typhi CT18.;"				
RL	Nature 413:848-852(2001).				
RN	[4]				
RP	SEQUENCE FROM N.A.				
RC	SPECIES=S.typhi; STRAIN=Ty2 / ATCC 700931;				
RX	MEDLINE=22531357; PubMed=12644504;				
RA	Deng W., Liou S.-R., Plunkett G. III, Mayhew G.F., Rose D.J.,				
RA	Burland V., Kodoyanni V., Schwartz D.C., Blattner P.R.;				

RT "Comparative genomics of Salmonella enterica serovar Typhi strains Ty2  
and CT18.";  
RL J. Bacteriol. 185:2330-2337(2003).  
RN [5]

RP SEQUENCE FROM N.A.  
RC SPECIES=S.enteritidis; STRAIN=27655-3B;  
RX MEDLINE=96146512; PubMed=8550497;  
RA Collinson S.K., Clouthier S.C., Doran J.L., Baner P.A., Kay W.W.;  
RT "Salmonella enteritidis agfBAC operon encoding thin, aggregative  
fimbriae.";  
RL J. Bacteriol. 178:662-667(1996).  
RN [6]

RP SEQUENCE OF 21-151 FROM N.A.  
RC SPECIES=S.enteritidis; STRAIN=27655-3B;  
RX MEDLINE=94013373; PubMed=8104955;  
RA Doran J.L., Collinson S.K., Burian J., Sarlos G., Todd E.C.D.,  
RA Munro C.K., Kay C.M., Baner P.A., Peterkin P.I., Kay W.W.;  
RT "DNA-based diagnostic tests for Salmonella species targeting agfA,  
the structural gene for thin, aggregative fimbriae.";  
RL J. Clin. Microbiol. 31:2263-2273(1993).  
RN [7]

RP SEQUENCE OF 21-33.  
RC SPECIES=S.enteritidis; STRAIN=27655-3B;  
RX MEDLINE=91310586; PubMed=1677357;  
RA Collinson S.K., Emeody L., Mueller K.M., Trust T.J., Kay W.W.;  
RT "Purification and characterization of thin, aggregative fimbriae from  
Salmonella enteritidis.";  
RL J. Bacteriol. 173:4773-4781(1991).

CC -!- FUNCTION: CURLIN IS THE STRUCTURAL SUBUNIT OF THE CURLI. CURLI ARE  
CC COILED SURFACE STRUCTURES THAT ASSEMBLE PREFERENTIALLY AT GROWTH  
CC TEMPERATURES BELOW 37 DEGREES CELSIUS. CURLI CAN BIND TO  
CC FIBRONECTIN.

CC -!- SIMILARITY: BELONGS TO THE CSGA/CSGB FAMILY.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

DR EMBL; AJ002301; CAA05317.1; -;  
DR EMBL; AE008749; AAL20074.1; -;  
DR EMBL; AL627269; CAD08268.1; -;  
DR EMBL; AE016840; AAO69399.1; -;  
DR EMBL; U43280; AAC43599.1; -;  
DR PIR; JC6039; JC6039.  
DR StyGene; SG10608; csGA.  
KW Fimbria; Signal; Complete proteome.  
FT SIGNAL 1 20  
FT CHAIN 21 151 MAJOR CURLIN SUBUNIT.  
FT CONFLICT 134 151 SVMRVQVGFNNATANYQ -> DSTVQVAS (IN  
FT REF. 6).  
SQ SEQUENCE 151 AA; 15305 MW; B7DAC0D16B621359 CRC64;

Query Match 88.7%; Score 681; DB 1; Length 151;  
Best Local Similarity 90.7%; Pred. No. 1.8e-50;  
Matches 137; Conservative 2; Mismatches 12; Indels 0; Gaps 0;

QY 1 MKLLKVAFAAIVVSGSALAGVVPQWGGNGNNGGNSGGPDSTLSIYQYGSANAALALQ 60  
Db 1 MKLLKVAFAAIVVSGSALAGVVPQWGGNGNNGGNSGGPDSTLSIYQYGSANAALALQ 60

QY 61 SPARKSETTITGSGYNGADVGGADNSTIETQNGFRNNATYDQLVTRVVTHEMAHAGG 120  
Db 61 SPARKSETTITGSGYNGADVGGADNSTIETQNGFRNNATYDQLVTRVVTHEMAHAGG 120

QY 121 NNAALVNQASDSSVMRVQVGFNNATANYQ 151  
Db 121 NNAALVNQASDSSVMRVQVGFNNATANYQ 151

RESULT 2  
ID CSGA\_ECOLI STANDARD; PRT; 151 AA.  
AC P28307;  
DT 01-DEC-1992 (Rel. 24, Created)  
DT 01-OCT-1996 (Rel. 34, Last sequence update)  
DT 28-FEB-2003 (Rel. 41, Last annotation update)  
DE Major curlin subunit precursor.  
GN CSGA OR B1042.  
OS Escherichia coli.  
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;  
OC Enterobacteriaceae; Escherichia.  
OX NCBI TaxID=562;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=K12 / W3110;  
RX MEDLINE=93211294; PubMed=8459772;  
RA Olsen A., Arngqvist A.;  
RT "The rpoS sigma factor relieves H-NS-mediated transcriptional  
repression of csGA, the subunit gene of fibronectin-binding curli in  
Escherichia coli.";  
RL Mol. Microbiol. 7:523-536(1993).  
RN [2]  
RP SEQUENCE FROM N.A.  
RC STRAIN=K12 / MC4100;  
RX MEDLINE=96414468; PubMed=8817489;  
RA Hammar M., Arngqvist A., Bian Z., Olsen A., Normark S.;  
RT "Expression of two csG operons is required for production of  
fibronectin- and congo red-binding curli polymers in Escherichia coli  
K-12.";  
RL Mol. Microbiol. 18:661-670(1995).  
RN [3]  
RP SEQUENCE FROM N.A.  
RC STRAIN=K12 / MG1655;  
RX MEDLINE=97426617; PubMed=9278503;  
RA Blatter F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,  
RA Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,  
RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,  
RA Mau B., Shao Y.;  
RT "The complete genome sequence of Escherichia coli K-12.";  
RL Science 277:1453-1474(1997).  
RN [4]  
RP SEQUENCE FROM N.A.  
RC STRAIN=K12;  
RX MEDLINE=97061202; PubMed=8905232;  
RA Oshima T., Aiba H., Baba T., Fujita K., Hayashi K., Honjo A.,  
RA Ikemoto K., Inada T., Itoh T., Kajihara M., Kanai K., Kashimoto K.,  
RA Kimura S., Kitagawa M., Makino K., Masuda S., Miki T., Mizobuchi K.,  
RA Mori H., Motomura K., Nakamura Y., Nishimoto H., Nishio Y., Saito N.,  
RA Sampei G., Seki Y., Tagami H., Takemoto K., Wada C., Yamamoto Y.,  
RA Yano M., Horiuchi T.;  
RT "A 718-Kb DNA sequence of the Escherichia coli K-12 genome  
corresponding to the 12.7-28.0 min region on the linkage map.";  
RL DNA Res. 3:137-155(1996).  
RN [5]  
RP SEQUENCE OF 21-40.  
RC STRAIN=K12 / YMEI;  
RX MEDLINE=93033873; PubMed=1357528;  
RA Arngqvist A., Olsen A., Pfeifer J., Russell D.G., Normark S.;  
RT "The Crl protein activates cryptic genes for curli formation and  
fibronectin binding in Escherichia coli HB101.";  
RL Mol. Microbiol. 6:2443-2452(1992).  
RN [6]  
RP SEQUENCE OF 21-31.  
RC MEDLINE=91310586; PubMed=1677357;  
RA Collinson S.K., Emeody L., Trust T.J., Kay W.W.;  
RT "Purification and characterization of thin, aggregative fimbriae from  
Salmonella enteritidis.";  
RL J. Bacteriol. 173:4773-4781(1991).  
CC -!- FUNCTION: CURLIN IS THE STRUCTURAL SUBUNIT OF THE CURLI. CURLI ARE  
CC COILED SURFACE STRUCTURES THAT ASSEMBLE PREFERENTIALLY AT GROWTH  
CC TEMPERATURES BELOW 37 DEGREES CELSIUS. CURLI CAN BIND TO  
CC FIBRONECTIN.

```

CC  -!- SIMILARITY: BELONGS TO THE CSGA/CSGB FAMILY.
CC  -----
CC  This SWISS-PROT entry is copyright. It is produced through a collaboration
CC  between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC  the European Bioinformatics Institute. There are no restrictions on its
CC  use by non-profit institutions as long as its content is in no way
CC  modified and this statement is not removed. Usage by and for commercial
CC  entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC  or send an email to license@isb-sib.ch).
CC  -----
CC  EMBL; L04979; AAA23616.1; -
CC  EMBL; X90754; CAAG2282.1; -
CC  EMBL; AB000205; AAC74126.1; -
CC  EMBL; D90741; BAA35832.1; -
CC  EMBL; D90742; BAA35840.1; -
CC  PIR; S70788; S70788.
CC  ECoGene; EGI1489; csGA.
CC  Fimbrria; Signal; Complete proteome.
CC  SIGNAL 1 20 MAJOR CURLIN SUBUNIT.
CC  CHAIN 21 151 A -> E (IN REF. 1).
CC  CONFLICT 7 7
CC  SEQUENCE 151 AA; 15049 MW; C003470D208D395F CRC64;
CC  -----
CC  Query Match 67.3%; Score 517; DB 1; Length 151;
CC  Best Local Similarity 69.5%; Pred. No. 9.7e-37;
CC  Matches 105; Conservative 17; Mismatches 29; Indels 0; Gaps 0;
CC  -----
CC  QY 1 MKLLKVAAPAAIVVSGSALAGVVPQWGGGNGGNGSGSPDSTLSIYQYGSANAALALQ 60
CC  DB 1 MKLLKVAAPAAIVVSGSALAGVVPQWGGGNGGNGSGPNSLNIYQYGGGNSALALQ 60
CC  QY 61 SDARKSETTTTQSGYNGADVGQGDADNSTIETQNGFRNATYDQVLTWVTHEMAHAGG 120
CC  DB 61 TDARNSLTITQGGNGADVGQGDSSIDLTRQFGNSATLTDQWNGKSNEMTVKQFGG 120
CC  QY 121 NNAALVNOTASDSSVMVYQVFGNNATANYQ 151
CC  DB 121 GNGAAVDQTASNSVNTVQVFGNNATAHQY 151
CC  -----
CC  RESULT 3
CC  CSGA_ECO57
CC  ID CSGA_ECO57 STANDARD; PRT; 152 AA.
CC  AC Q93U24;
CC  DT 28-FEB-2003 (Rel. 41, Created)
CC  DT 28-FEB-2003 (Rel. 41, Last sequence update)
CC  DT 28-FEB-2003 (Rel. 41, Last annotation update)
CC  DE Major curlin subunit precursor.
CC  GN CSGA OR Z1676 OR ECS1420.
CC  OS Escherichia coli O157:H7.
CC  OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
CC  OC Enterobacteriaceae; Escherichia.
CC  OX NCBI_TaxID=83334;
CC  RN [1]
CC  RP SEQUENCE FROM N.A.
CC  RC STRAIN=O157:H7 / ATCC 43895;
CC  RX MEDLINE=21218556; PubMed=11319125;
CC  RA Unlich G.A., Keen J.E., Elder R.O.;
CC  RT "Mutations in the csGP promoter associated with variations in curli
CC  expression in certain strains of Escherichia coli O157:H7.";
CC  RL Appl. Environ. Microbiol. 67:2367-2370(2001).
CC  [2]
CC  RP SEQUENCE FROM N.A.
CC  RC STRAIN=O157:H7 / EDL933 / ATCC 700927;
CC  RX MEDLINE=21074935; PubMed=11206551;
CC  RA Perna N.T., Plunkett G. III, Burland V., Mau B., Glasner J.D.,
CC  RA Rose D.J., Mayhew G.F., Evans P.S., Gregor J., Kirkpatrick H.A.,
CC  RA Posfai G., Hackett J., Klink S., Boutin A., Shao Y., Miller L.,
CC  RA Grotbeck E.J., Davis N.W., Lim A., Dimalanta E.T., Potanousis K.,
CC  RA Apodaca J., Anantharaman T.S., Lin J., Yen G., Schwartz D.C.,
CC  RA Welch R.A., Blattner F.R.;
CC  RT "Genome sequence of enterohaemorrhagic Escherichia coli O157:H7.";
CC  RL Nature 409:529-533(2001).

```

```

RN  [3]
RP  SEQUENCE FROM N.A.
RX  STRAIN=O157:H7 / RMD 0509952;
RX  MEDLINE=21156231; PubMed=11258796;
RA  Hayashi T., Makino K., Ohnishi M., Kurokawa K., Ishii K., Yokoyama K.,
RA  Han C.-G., Ohtsubo E., Nakayama K., Murata T., Tanaka M., Tobe T.,
RA  Iida T., Takami H., Honda T., Sasakawa C., Ogasawara N., Yasunaga T.,
RA  Kuhara S., Shiba T., Hattori M., Shinagawa H.;
RT  "Complete genome sequence of enterohaemorrhagic Escherichia coli
RT  O157:H7 and genomic comparison with a laboratory strain K-12.";
RL  DNA Res. 8:11-22(2001).
CC  -!- FUNCTION: CURLIN IS THE STRUCTURAL SUBUNIT OF THE CURLI. CURLI ARE
CC  COILED SURFACE STRUCTURES THAT ASSEMBLE PREFERENTIALLY AT GROWTH
CC  TEMPERATURES BELOW 37 DEGREES CELSIUS. CURLI CAN BIND TO
CC  FIBRONECTIN.
CC  -!- SIMILARITY: BELONGS TO THE CSGA/CSGB FAMILY.
CC  -----
CC  This SWISS-PROT entry is copyright. It is produced through a collaboration
CC  between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC  the European Bioinformatics Institute. There are no restrictions on its
CC  use by non-profit institutions as long as its content is in no way
CC  modified and this statement is not removed. Usage by and for commercial
CC  entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC  or send an email to license@isb-sib.ch).
CC  -----
CC  EMBL; AF275733; AAK53212.1; -
CC  EMBL; AE005315; AAG55788.1; -
CC  EMBL; AP002554; BAB34843.1; -
CC  PIR; D90806; D90806.
CC  PIR; H85685; H85685.
CC  Fimbrria; Signal; Complete proteome.
CC  SIGNAL 1 20 BY SIMILARITY.
CC  CHAIN 21 152 MAJOR CURLIN SUBUNIT.
CC  SEQUENCE 152 AA; 15099 MW; EE2D2D94DDE91243 CRC64;
CC  -----
CC  Query Match 65.4%; Score 502.5; DB 1; Length 152;
CC  Best Local Similarity 68.4%; Pred. No. 1.6e-35;
CC  Matches 104; Conservative 18; Mismatches 29; Indels 1; Gaps 1;
CC  -----
CC  QY 1 MKLLKVAAPAAIVVSGSALAGVVPQW-GGGNGHNGGNGSGPDSSTLSIYQYGSANAALAL 59
CC  DB 1 MKLLKVAAPAAIVVSGSALAGVVPQWGGGNGGNGSGPNSLNIYQYGGGNSALAL 60
CC  QY 60 QSDARKSETTTTQSGYNGADVGQGDADNSTIETQNGFRNATYDQVLTWVTHEMAHAG 119
CC  DB 61 QADARNSLTITQGGNGADVGQGDSSIDLTRQFGNSATLTDQWNGKSHMTVKQFG 120
CC  QY 120 GNNALVNOTASDSSVMVYQVFGNNATANYQ 151
CC  DB 121 GNGAAVDQTASNSVNTVQVFGNNATAHQY 152
CC  -----
CC  RESULT 4
CC  CSGB_ECOLI
CC  ID CSGB_ECOLI STANDARD; PRT; 151 AA.
CC  AC P39828;
CC  DT 01-FEB-1995 (Rel. 31, Created)
CC  DT 01-OCT-1996 (Rel. 34, Last sequence update)
CC  DT 28-FEB-2003 (Rel. 41, Last annotation update)
CC  DE Minor curlin subunit precursor.
CC  GN CSGB OR B1041 OR Z1675 OR ECS1419.
CC  OS Escherichia coli, and
CC  OS Escherichia coli O157:H7.
CC  OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
CC  OC Enterobacteriaceae; Escherichia.
CC  OX NCBI_TaxID=562, 83334;
CC  RN [1]
CC  RP SEQUENCE FROM N.A.
CC  RC STRAIN=K12 / MC4100;
CC  RX MEDLINE=9641468; PubMed=8817489;
CC  RA Hammar M., Arngqvist A., Bian Z., Olsen A., Normark S.;
CC  RT "Expression of two csg operons is required for production of
CC  fibronectin- and congo red-binding curli polymers in Escherichia coli
CC  RT

```

```

RT K-12."
RL Mol. Microbiol. 18:661-670(1995).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=K12 / MGI655;
RX MEDLINE=97426617; PubMed=9278503;
RA Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,
RA Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,
RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,
RA Mau B., Shao Y.;
RT "The complete genome sequence of Escherichia coli K-12."
RL Science 277:1453-1474(1997).
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=K12;
RX MEDLINE=97061202; PubMed=8905232;
RA Oshima T., Aiba H., Baba T., Fujita K., Hayashi K., Honjo A.,
RA Ikemoto K., Inada T., Itoh T., Kajihara M., Kanai K., Kashimoto K.,
RA Kimura S., Kitagawa M., Makino K., Masuda S., Miki T., Mizobuchi K.,
RA Mori H., Motomura K., Nakamura Y., Nishimoto H., Nishio Y., Saito N.,
RA Sampei G., Seki Y., Tagami H., Takemoto K., Wada C., Yamamoto Y.,
RA Yano M., Horiuchi T.;
RT "A 718-kb DNA sequence of the Escherichia coli K-12 genome
RT corresponding to the 12.7-28.0 min region on the linkage map."
RL DNA Res. 3:137-155(1996).
RN [4]
RP SEQUENCE FROM N.A.
RC STRAIN=O157:H7 / EDL933 / ATCC 700927;
RX MEDLINE=21074935; PubMed=11206551;
RA Perna N.T., Plunkett G. III, Burland V., Mau B., Glasner J.D.,
RA Rose D.J., Mayhew G.F., Evans P.S., Gregor J., Kirkpatrick H.A.,
RA Posfai G., Hackett J., Klink S., Boutin A., Shao Y., Miller L.,
RA Grobeck E.J., Davis N.W., Lim A., Dimalanta E.T., Potamousis K.,
RA Apodaca J., Anantharaman T.S., Lin J., Yen G., Schwartz D.C.,
RA Welch R.A., Blattner F.R.;
RT "Genome sequence of enterohaemorrhagic Escherichia coli O157:H7."
RL Nature 409:529-533(2001).
RN [5]
RP SEQUENCE FROM N.A.
RC STRAIN=O157:H7 / RIMD 0509952;
RX MEDLINE=21156231; PubMed=11258796;
RA Hayashi T., Makino K., Ohnishi M., Kurokawa K., Ishii K., Yokoyama K.,
RA Han C.-G., Ohtsubo E., Nakayama K., Murata T., Tanaka M., Tobe T.,
RA Iida T., Takami H., Honda T., Sasakawa C., Ogasawara N., Yasunaga T.,
RA Kuhara S., Shiba T., Rattori M., Shinagawa H.;
RT "Complete genome sequence of enterohemorrhagic Escherichia coli
RT O157:H7 and genomic comparison with a laboratory strain K-12."
RL DNA Res. 8:11-22(2001).
RN [6]
RP SEQUENCE OF 1-21 FROM N.A.
RC STRAIN=K12;
RX MEDLINE=95157246; PubMed=7854117;
RA Arngvist A., Olsen A., Normark S.;
RT "Sigma S-dependent growth-phase induction of the csfBA promoter in
RT Escherichia coli can be achieved in vivo by sigma 70 in the absence
RT of the nucleoid-associated protein H-NS."
RL Mol. Microbiol. 13:1021-1032(1994).
CC -1- FUNCTION: CURLIN IS THE STRUCTURAL SUBUNIT OF THE CURLI. CURLI ARE
CC COILED SURFACE STRUCTURES THAT ASSEMBLE PREFERENTIALLY AT GROWTH
CC TEMPERATURES BELOW 37 DEGREES CELSIUS. CURLI CAN BIND TO
CC FIBRONECTIN. THE MINOR SUBUNIT IS THE NUCLEATION COMPONENT OF
CC CURLIN MONOMERS.
CC -1- SIMILARITY: BELONGS TO THE CSGA/CSGB FAMILY.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC EMBL; X90754; CAA62281.1;

```

```

DR EMBL; AE000205; AAC74125.1; -.
DR EMBL; D90741; BAA35831.1; -.
DR EMBL; AE005315; AAG55787.1; -.
DR EMBL; AP002554; BAB34842.1; -.
DR PIR; C90806; C90806.
DR PIR; G85665; G85665.
DR PIR; S70787; S70787.
DR EcoGene; EGI12621; csfB.
KW Fibria; Signal; Complete proteome.
FT SIGNAL 1 21 POTENTIAL.
FT CHAIN 22 151 MINOR CURLIN SUBUNIT.
SQ SEQUENCE 151 AA; 15882 MW; B18D266B964014B8 CRC64;

Query Match 13.6%; Score 104.5; DB 1; Length 151;
Best Local Similarity 29.1%; Pred. No. 0.033;
Matches 34; Conservative 15; Mismatches 57; Indels 11; Gaps 3;

Qy 38 SSGPSTLSIYQYGSANAALQSDARKSETTTTQSGYNGADVGOGADNSTIELTQNGF 97
Db 21 AAGYDLANSEYNF----AVNELSKSFNQAIICQAGTNNSAQLRQGGKLLAVVAQEGS 76
Qy 98 RNNATYDQLVTRVVTHEMAH---AGGNNALVNAQTASDSSVMVRQVGFGNATANQY 151
Db 77 SNRAKIDQ----TGDYNLAVIDQAGSANDASISQAGYNTAMIIQKSGNKANITQY 129

RESULT 5
CSGB SALT1 STANDARD; PRT; 151 AA.
AC Q827M3;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Minor curlin subunit precursor.
GN CSGB OR STY1180 OR T1777.
OS Salmonella typhi.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Salmonella.
OC NCBI_TaxID=601;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CT18;
RX MEDLINE=21534947; PubMed=11677608;
RA Parkhill J., Dougan G., James K.D., Thomson N.R., Pickard D., Wain J.,
RA Churcher C., Mungall K.L., Bentley S.D., Holden M.T.G., Sebahia M.,
RA Baker S., Basham D., Brooks K., Chillingworth T., Connerton P.,
RA Cronin A., Davis P., Davies R.M., Dowd L., White N., Farrar J.,
RA Feltwell T., Hamlin N., Haque A., Hien T.T., Holroyd S., Jagers K.,
RA Krogh A., Larsen T.S., Leather S., Moule S., O'Gaora P., Parry C.,
RA Quail M.A., Rutherford K., Simmonds M., Skelton J., Stevens K.,
RA Whitehead S., Barrett B.G.;
RT "Complete genome sequence of a multiple drug resistant Salmonella
RT enterica serovar Typhi CT18."
RL Nature 413:848-852(2001).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=Ty2 / ATCC 700931;
RX MEDLINE=22531367; PubMed=12644504;
RA Deng W., Liou S.-R., Plunkett G. III, Mayhew G.F., Rose D.J.,
RA Burland V., Kodyanni V., Schwartz D.C., Blattner F.R.;
RT "Comparative genomics of Salmonella enterica serovar Typhi strains Ty2
RT and CT18."
RL J. Bacteriol. 185:2330-2337(2003).
CC -1- FUNCTION: CURLIN IS THE STRUCTURAL SUBUNIT OF THE CURLI. CURLI ARE
CC COILED SURFACE STRUCTURES THAT ASSEMBLE PREFERENTIALLY AT GROWTH
CC TEMPERATURES BELOW 37 DEGREES CELSIUS. CURLI CAN BIND TO
CC FIBRONECTIN. THE MINOR SUBUNIT IS THE NUCLEATION COMPONENT OF
CC CURLIN MONOMERS.
CC -1- SIMILARITY: BELONGS TO THE CSGA/CSGB FAMILY.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its

```

use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

EMBL; AL627269; CAD08267.1; -  
 DR EMBL; AE016840; AAO69400.1; -  
 KW Fimbria; Signal; Complete proteome.  
 FT SIGNAL 1 21 POTENTIAL.  
 FT CHAIN 22 151 MINOR CURLIN SUBUNIT.  
 SQ SEQUENCE 151 AA; 16254 MW; 161C54326E573495 CRC64;

Query Match 13.0%; Score 99.5; DB 1; Length 151;  
 Best Local Similarity 28.2%; Pred. No. 0.088;  
 Matches 33; Conservative 16; Mismatches 51; Indels 17; Gaps 4;

QY 51 GSANAALALQSDARKSE-----TTITQSGYNGADVGO-CADNST-----IELTQ 94

Db 14 GAGIATATNYDLARSEYNFAVNELSKSFNQAAIIGQVGTDSARVREGSKLLSVISQ 73

QY 95 NGFRNNATYDQLVTRVVTHEMAHAGGNAALVNQTASDSVMVRQVGFGNATANOY 151

Db 74 EGENNRKAVDQAGNYNFAY-IEQTGNANDASISQSAVGNAAIIQKSGNKANITQY 129

#### RESULT 6

CSGB\_SALTY STANDARD; PRT; 151 AA.

AC P55226;

DT 01-OCT-1996 (Rel. 34, Created)

DT 01-OCT-1996 (Rel. 34, Last sequence update)

DT 28-FEB-2003 (Rel. 41, Last annotation update)

DE Minor curlin subunit precursor (Fimbrin SEF17 minor subunit).

GN CSGB OR AGFB OR STM1143.

OS Salmonella typhimurium, and

OS Salmonella enteritidis.

OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;

OC Enterobacteriaceae; Salmonella.

OX NCBI TaxID=602, 592;

EN [1]

RP SEQUENCE FROM N.A.

RC SPECIES=S.typhimurium; STRAIN=SR-11;

RA MEDLINE=98117058; PubMed=9457980;

RA Romling U., Bian Z., Hammar M., Sierralta W.D., Normark S.;

RT "Curli fibers are highly conserved between Salmonella typhimurium and

Escherichia coli with respect to operon structure and regulation.";

RL J. Bacteriol. 180:722-731(1998).

EN [2]

RP SEQUENCE FROM N.A.

RC SPECIES=S.typhimurium; STRAIN=LT2 / SGSC1412 / ATCC 700720;

RA MEDLINE=21534948; PubMed=11677609;

RA McClelland M., Sanderson K.E., Spieth J., Clifton S.W., Latreille P.,

Courtney L., Porwollik S., Ali J., Dante M., Du F., Hou S., Layman D.,

Leonard S., Nguyen C., Scott K., Holmes A., Grewal N., Mulvaney E.,

Ryan E., Sun H., Florea L., Miller W., Stoneking T., Nhan M.,

Waterston R., Wilson R.K.;

RT "Complete genome sequence of Salmonella enterica serovar Typhimurium

LT2.";

RL Nature 413:852-856(2001).

EN [3]

RP SEQUENCE FROM N.A.

RC SPECIES=S.enteritidis; STRAIN=27655-3B;

RA MEDLINE=96146512; PubMed=8550497;

RA Collinson S.K., Clouthier S.C., Doran J.L., Baner P.A., Kay W.W.;

RT "Salmonella enteritidis afBAC operon encoding thin, aggregative

fimbriae.";

RL J. Bacteriol. 178:662-667(1996).

CC -!- FUNCTION: CURLIN IS THE STRUCTURAL SUBUNIT OF THE CURLI. CURLI ARE

COILED SURFACE STRUCTURES THAT ASSEMBLE PREFERENTIALLY AT GROWTH

TEMPERATURES BELOW 37 DEGREES CELSIUS. CURLI CAN BIND TO

FIBRONECTIN. THE MINOR SUBUNIT IS THE NUCLEATION COMPONENT OF

CURLIN MONOMERS.

CC -!- SIMILARITY: BELONGS TO THE CSGB/CSGB FAMILY.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

DR EMBL; AJ002301; CAA05316.1; -  
 DR EMBL; AE008749; AAL20073.1; -  
 DR EMBL; U43280; AAC43598.1; -  
 DR PIR; JC6040; JC6040.  
 DR StyGene; SG10609; CSG.

KW Fimbria; Signal; Complete proteome.

FT SIGNAL 1 21 POTENTIAL.

FT CHAIN 22 151 MINOR CURLIN SUBUNIT.

SQ SEQUENCE 151 AA; 16182 MW; C0FC5430E6DD361D CRC64;

Query Match 13.0%; Score 99.5; DB 1; Length 151;  
 Best Local Similarity 28.2%; Pred. No. 0.088;  
 Matches 33; Conservative 16; Mismatches 51; Indels 17; Gaps 4;

QY 51 GSANAALALQSDARKSE-----TTITQSGYNGADVGO-CADNST-----IELTQ 94

Db 14 GAGIATATNYDLARSEYNFAVNELSKSFNQAAIIGQVGTDSARVREGSKLLSVISQ 73

QY 95 NGFRNNATYDQLVTRVVTHEMAHAGGNAALVNQTASDSVMVRQVGFGNATANOY 151

Db 74 EGENNRKAVDQAGNYNFAY-IEQTGNANDASISQSAVGNAAIIQKSGNKANITQY 129

#### RESULT 7

OVO\_DROME

ID\_OVO\_DROME STANDARD; PRT; 1028 AA.

AC P51521; Q9XZU4;

DT 01-OCT-1996 (Rel. 34, Created)

DT 01-OCT-1996 (Rel. 34, Last sequence update)

DT 10-OCT-2003 (Rel. 42, Last annotation update)

DE OVO protein (Shaven baby protein).

GN OVO OR SVB.

OS Drosophila melanogaster (Fruit fly).

OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;

OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;

OC Ephydroidea; Drosophilidae; Drosophila.

OX NCBI TaxID=7227;

EN [1]

RP SEQUENCE FROM N.A.

RC TISSUE=Ovary;

RA MEDLINE=95021209; PubMed=7935398;

RA Garfinkel M.D., Wang J., Liang Y., Mahowald A.P.;

RT "Multiple products from the shavenbaby-ovo gene region of Drosophila

melanogaster: relationship to genetic complexity.";

RL Mol. Cell. Biol. 14:6809-6818(1994).

EN [2]

RP SEQUENCE FROM N.A.

RC STRAIN=Oregon-R;

RA MEDLINE=91293102; PubMed=1712294;

RA Mevel-Ninio M.T.M., Terracol R., Kafatos F.C.;

RT "The ovo gene of Drosophila encodes a zinc finger protein required

for female germ line development.";

RL EMO J. 10:2259-2266(1991).

CC -!- FUNCTION: REQUIRED FOR SURVIVAL AND DIFFERENTIATION OF FEMALE GERM

CELLS. PLAYS A ROLE IN GERM LINE SEX DETERMINATION.

CC -!- SUBCELLULAR LOCATION: Nuclear (potential).

CC -!- DEVELOPMENTAL STAGE: FIRST APPEARS IN THE GERMARINUM AND

ACCUMULATES IN NURSE CELLS DURING OÖGENESIS. STORED IN THE EGG,

BUT IS RAPIDLY LOST IN THE EMBRYOS EXCEPT FOR ITS CONTINUED

PRESENCE IN THE GERM LINE PRECURSOR POLE CELLS.

CC -!- SIMILARITY: Contains 4 C2H2-type zinc fingers.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration

between the Swiss Institute of Bioinformatics and the EMBL outstation -

CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----

DR EMBL; U11383; AAB60216.1; -;  
 DR EMBL; X59772; CAB36921.1; ALT\_SEQ.  
 DR PIR; A56038; A56038.  
 DR HSP; P07248; 2ADR.  
 DR TRANSFAC; T00669; -;  
 DR FlyBase; FBgn0003028; ovo.  
 DR InterPro; IPR007087; znf\_C2H2.  
 DR Pfam; PF00096; zf-C2H2; 3.  
 DR SMART; SM00355; Znf\_C2H2; 4.  
 DR PROSITE; PS00028; ZINC\_FINGER\_C2H2\_1; 3.  
 DR PROSITE; PS0157; ZINC\_FINGER\_C2H2\_2; 3.  
 DR Zinc-finger; Metal-binding; DNA-binding; Repeat; Nuclear protein;  
 KW Transcription regulation.  
 FT DOMAIN 62 65 POLY-ALA.  
 FT DOMAIN 72 77 POLY-GLY.  
 FT DOMAIN 80 85 POLY-GLY.  
 FT DOMAIN 98 108 POLY-GLY.  
 FT DOMAIN 144 152 POLY-HIS.  
 FT DOMAIN 153 159 POLY-ASN.  
 FT DOMAIN 336 339 POLY-GLN.  
 FT DOMAIN 347 353 POLY-GLN.  
 FT DOMAIN 357 361 POLY-GLN.  
 FT DOMAIN 410 414 POLY-GLN.  
 FT DOMAIN 418 422 POLY-GLN.  
 FT DOMAIN 426 432 POLY-GLN.  
 FT DOMAIN 445 453 POLY-GLN.  
 FT DOMAIN 456 459 POLY-GLN.  
 FT DOMAIN 466 474 POLY-GLN.  
 FT DOMAIN 497 517 POLY-ALA.  
 FT DOMAIN 524 529 POLY-SER.  
 FT DOMAIN 549 558 POLY-ALA.  
 FT DOMAIN 639 651 POLY-ALA.  
 FT DOMAIN 717 725 POLY-ALA.  
 FT DOMAIN 797 802 POLY-GLN.  
 FT DOMAIN 820 823 POLY-GLN.  
 FT DOMAIN 826 832 POLY-GLN.  
 FT ZN\_FING 874 896 C2H2-TYPE 1.  
 FT ZN\_FING 902 924 C2H2-TYPE 2.  
 FT ZN\_FING 930 953 C2H2-TYPE 3.  
 FT ZN\_FING 969 992 C2H2-TYPE 4.  
 FT CONFLICT 647 647 A -> R (IN REF. 2).  
 SQ SEQUENCE 1028 AA; 110620 MW; D7068BB2BC0F6F77 CRC64;  
 Query Match 12.0%; Score 92.5; DB 1; Length 1028;  
 Best Local Similarity 26.0%; Pred. No. 2.9;  
 Matches 38; Conservative 15; Mismatches 66; Indels 27; Gaps 4;  
 QY 3 LLKVAFAAIVVSGSALACVWPQWGGGNGHNGGNSGPDSTLSIYQVGSANAALALQSD 62  
 Db 59 LQNAFAAIVVIMAGSG-----GGCTGNGGGASGPGGPGSANGGGGG----- 104  
 QY 63 ARKSETTITQSGYNGADYVGQAGDINSTIELTQNGFRNNATYQLVTRVVVTHMAHAGNN 122  
 Db 105 -----GGGYINGCGVG-GPNNSLDGNLLNFASVSNYESKFKHHHHHHQHN 155  
 QY 123 AALVNOTASDSSVMVQVQFGNNATA 148  
 Db 156 -----NNNGGQTSMMGHPPYGNPSA 177

RESULT 8

ID TNK1 HUMAN STANDARD; PRT; 1327 AA.  
 AC Q95271; Q95272;  
 DT 28-FEB-2003 (Rel. 41, Created)  
 DT 28-FEB-2003 (Rel. 41, Last sequence update)  
 DT 10-OCT-2003 (Rel. 42, Last annotation update)

DE Tankyrase 1 (EC 2.4.2.30) (TANK1) (Tankyrase I) (TNKS-1) (TRF1-  
 DE interacting ankyrin-related ADP-ribose polymerase).  
 OS TNKS OR TNKS1 OR TINI OR TINF1 OR PARPL.  
 OS Homo sapiens (Human).  
 CC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 CC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 CC NCBL\_taxid=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A. (ISOFORMS 1 AND 2).  
 RC TISSUE=Testis;  
 RX MEDLINE=99040105; PubMed=9822378;  
 RA Smith S., Giriat I., Schmitt A., de Lange T.;  
 RT "Tankyrase, a poly(ADP-ribose) polymerase at human telomeres.";  
 RL Science 282:1484-1487(1998).  
 RN [2]  
 RP SUBCELLULAR LOCATION.  
 RX MEDLINE=99454782; PubMed=10523501;  
 RA Smith S., de Lange T.;  
 RT "Cell cycle dependent localization of the telomeric PARP, tankyrase,  
 RT to nuclear pore complexes and centrosomes.";  
 RL J. Cell Sci. 112:3649-3656(1999).  
 RN [3]  
 RP FUNCTION, AND PHOSPHORYLATION.  
 RX MEDLINE=20556282; PubMed=10988299;  
 RA Chi N.-W., Iodish H.F.;  
 RT "Tankyrase is a Golgi-associated mitogen-activated protein kinase  
 RT substrate that interacts with IRAP in GLUT4 vesicles.";  
 RL J. Biol. Chem. 275:38437-38444(2000).  
 RN [4]  
 RP FUNCTION, AND MUTAGENESIS OF HIS-1184 AND GLU-1291.  
 RX MEDLINE=21602874; PubMed=11739745;  
 RA Cook B.D., Dynek J.N., Chang W., Shostak G., Smith S.;  
 RT "Role for the related poly(ADP-ribose) polymerases tankyrase 1 and 2  
 RT at human telomeres.";  
 RL Mol. Cell. Biol. 22:332-342(2002).  
 CC -!- FUNCTION: May regulate vesicle trafficking and modulate the  
 CC subcellular distribution of SLC2A4/GLUT4-vesicles. Has PARP  
 CC activity and can modify TRF1, and thereby contribute to the  
 CC regulation of telomere length.  
 CC -!- CATALYTIC ACTIVITY. NAD(+) + {ADP-D-ribose}[N]-acceptor =  
 CC nicotinamide + {ADP-D-ribose}[N+1]-acceptor.  
 CC -!- SUBUNIT: Oligomerizes and associates with TNKS2. Interacts with  
 CC the cytoplasmic domain of LNPEP/Otase in SLC2A4/GLUT4-vesicles.  
 CC Binds to the N-terminus of telomeric TRF1 via the ANK repeats.  
 CC -!- SUBCELLULAR LOCATION: Cytoplasmic, associated with the Golgi and  
 CC with juxtanuclear SLC2A4/GLUT4-vesicles. A minor proportion is  
 CC also found at nuclear pore complexes and around the pericentriolar  
 CC matrix of mitotic centrosomes. During interphase, a small fraction  
 CC of TNKS is found in the nucleus, associated with TRF1.  
 CC -!- ALTERNATIVE PRODUCTS:  
 CC Event=Alternative splicing; Named isoforms=2;  
 CC Name=1;  
 CC IsoId=O95271-1; Sequence=Displayed;  
 CC Name=2;  
 CC IsoId=O95271-2; Sequence=VSP\_004538, VSP\_004539;  
 CC Note=No experimental confirmation available;  
 CC -!- TISSUE SPECIFICITY: Ubiquitous; highest levels in testis.  
 CC -!- PTM: Upon insulin-stimulation, phosphorylated on serine residues  
 CC by MAPK kinases.  
 CC -!- PTM: ADP-ribosylated (-auto).  
 CC -!- SIMILARITY: Belongs to the PARP family.  
 CC -!- SIMILARITY: Contains 15 ANK repeats.  
 CC -!- SIMILARITY: Contains 1 sterile alpha motif (SAM) domain.  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 CC EMBL; AF082556; AAC79841.1; -;

```

DR EMBL; AF082557; AAC79842.1; -.
DR EMBL; AF082558; AAC79843.1; -.
DR EMBL; AF082559; AAC79844.1; -.
DR HSSP; Q00420; LAMC.
DR Genew; HGNC:11941; TNKS.
DR MIM; 603303; -.
DR GO; GO:0000781; C:chromosome, telomeric region; IDA.
DR GO; GO:0003950; F:NAD ADP-ribosyltransferase activity; IDA.
DR GO; GO:0005515; F:protein binding; IPI.
DR GO; GO:0007004; P:telomerase-dependent telomere maintenance; NAS.
DR InterPro; IPR002110; ANK.
DR InterPro; IPR001660; SAM.
DR Pfam; PF00023; ANK; 19.
DR Pfam; PF00536; SAM; 1.
DR PRINTS; PR01415; ANKYRIN.
DR SMART; SM00248; ANK; 17.
DR SMART; SM00454; SAM; 1.
DR PROSITE; PS00888; ANK REPEAT; 15.
DR PROSITE; PS0297; ANK REP REGION; 1.
DR PROSITE; PS0105; SAM DOMAIN; 1.
DR Transferrase; Glycosyltransferase; NAD; Golgi stack; Telomere;
KW Nuclear protein; Repeat; ANK repeat; ADP-ribosylation;
KW Phosphorylation; Alternative splicing.
FT REPEAT 215 247 ANK 1.
FT REPEAT 248 280 ANK 2.
FT REPEAT 281 313 ANK 3.
FT REPEAT 368 400 ANK 4.
FT REPEAT 401 433 ANK 5.
FT REPEAT 434 466 ANK 6.
FT REPEAT 521 556 ANK 7.
FT REPEAT 557 589 ANK 8.
FT REPEAT 590 622 ANK 9.
FT REPEAT 683 715 ANK 10.
FT REPEAT 716 748 ANK 11.
FT REPEAT 749 781 ANK 12.
FT REPEAT 836 868 ANK 13.
FT REPEAT 869 901 ANK 14.
FT REPEAT 902 934 ANK 15.
FT DOMAIN 1030 1089 SAM.
FT DOMAIN 1176 1327 PARP.
FT DOMAIN 9 14 POLY-HIS.
FT DOMAIN 27 34 POLY-PRO.
FT DOMAIN 128 134 POLY-SER.
FT DOMAIN 137 145 POLY-SER.
FT VARSPLIC 641 643
FT VARSPLIC 644 1327
FT MUTAGEN 1184 1184 WITH A-1291
FT MUTAGEN 1291 1291 E->A: LOSS OF ACTIVITY; WHEN ASSOCIATED
FT SEQUENCE 1327 AA; 142010 MW; E14DE985C710B957 CRC64;

Query Match 11.9%; Score 91.5; DB 1; Length 1327;
Best Local Similarity 29.4%; Pred. No. 4.7;
Matches 42; Conservative 17; Mismatches 65; Indels 19; Gaps 6;

QY 6 VAAPAAI-VVGSALAGVVPQWGGNHNHGGNSGDPSTLSIYQVGSANAALALQSDAR 64
Db 99 VAAAPVPAVSTSSAAGVAPNAGSGNSPSSSSPTSS-SSSSPSPGSSLAESPAA 157
QY 65 KSETTIT- - - - - QSGYGNADVGGADNSITELQNGFRNATYDQLVTRVWTHMAHAGG 120
Db 158 GVSSTAPLPGGAAGPGTGVPAVSGALNELLEACRNGD- - - - - VSRV- - - - - KRLVDAAN 206
QY 121 NNAALVNQTASDSSVMVRQVFG 143
Db 207 VNAK- - - - - DMAGKSSPLHFAAGFG 227

RESULT 9
OMP_RICUA STANDARD; PRT; 1656 AA.
ID OMP_RICUA STANDARD; PRT; 1656 AA.
AC O0653;
DT 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Outer membrane protein B precursor (168 kDa surface-layer protein)
DE (Surface protein antigen) (Cell surface antigen 5) (Scas) (rOmpB)
DE (rOmp B) [Contains: 120 kDa surface-exposed protein (Surface protein
DE antigen) (120 kDa outer membrane protein ompB); 32 kDa beta peptid].
GN OMPB
OS Rickettsia japonica.
OC Bacteria; Proteobacteria; Alphaproteobacteria; Rickettsiales;
OC Rickettsiaceae; Rickettsiae; Rickettsia.
OX NCBI_TaxID=35790;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-YH;
RA Uchiyama T.;
RT "Sequencing of the gene encoding the protein rOmp B of Rickettsia
RT japonica.";
RL Submitted (MAY-1997) to the EMBL/GenBank/DBJ databases.
CC -!- FUNCTION: THE 120 KDA SURFACE-EXPOSED PROTEIN IS A MAJOR
CC STRUCTURAL PROTEIN WHICH MAY PLAY A ROLE AS A RICKETTSIAL
CC VIRULENCE FACTOR AND/OR IMMUNOGEN DURING INFECTION (BY
CC similarity).
CC -!- FUNCTION: THE 32 KDA BETA PEPTIDE MAY SERVE AS A MEMBRANE ANCHOR
CC (By similarity).
CC -!- SUBCELLULAR LOCATION: Cell wall. This bacterium is covered by a S-
CC layer with hexagonal symmetry.
CC -!- SIMILARITY: BELONGS TO THE RICKETTSIAE OMPA/OMP FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; AB003681; BAA20138.1; -.
CC InterPro; IPR006315; Autotransport.
CC InterPro; IPR005546; Autotransporter.
CC Pfam; PF03797; Autotransporter; 1.
CC TIGRFAMs; TIGR01414; autotrans_bar1; 2.
KW Antigen; S-layer; Cell wall. 120 kDa SURFACE-EXPOSED PROTEIN.
FT CHAIN 1 1338 32 kDa BETA PEPTIDE.
FT DOMAIN 1339 1656 POLY-GLY.
FT SEQUENCE 1656 AA; 168097 MW; 3132A69C9DD5999F CRC64;

Query Match 11.7%; Score 90; DB 1; Length 1656;
Best Local Similarity 23.5%; Pred. No. 8;
Matches 40; Conservative 21; Mismatches 57; Indels 52; Gaps 6;

QY 6 VAAPAAI-VVGSALAGVVPQWGGNHNHGGNSGDPSTLSIYQVGSANAALALQSDAR 65
Db 509 VLAAGAITLDSATI- - - - - TGDIGNGG- - - - - GAALQSITLANDATK 547
QY 66 SETTITQSG- - - - - YNGADVQGGADNSITELT- - - - - QNGFRNATYD 104
Db 548 - - - - - TLTLGGANIISANGTINQANGGTTIKLTSTQNNIVVDCDLAIATDQTGVVDASLT 604
QY 105 QLVTRVWTHMAHAGGNAAL- - - - - VNQTASDSSVMVRQVFGNNAT 147
Db 605 NAQTITISGTIGIIGANNNTLTGQFNIGSKTTLNGGNVAINELVIGNGS 654

RESULT 10
N100 YEAST
ID N100 YEAST STANDARD; PRT; 959 AA.
AC Q02629;
DT 01-JUL-1993 (Rel. 26, Created)
DT 01-JUL-1993 (Rel. 26, Last sequence update)

```





DR EMBL; D89145; BAAL3807.1; -;  
 DR PIR; T50074; T50074.  
 DR HSP; P22629; ISWF.  
 DR GenDB SPombe; SPAC1486.05; -;  
 DR GO; GO:0016020; C:membrane; ISS.  
 DR GO; GO:0005643; C:nuclear pore; ISS.  
 DR GO; GO:0006606; P:protein-nucleus import; ISS.  
 DR InterPro; IPR007230; Nucleoporin2.  
 DR InterPro; IPR004325; Nucleoporin\_FG.  
 DR Pfam; PF04096; Nucleoporin2; 1.  
 DR Pfam; PF03093; Nucleoporin\_FG; 24.  
 DR Nucleolar protein; Transport; Repeat.  
 KW DOMAIN 9 625 GLY-RICH.  
 FT DOMAIN 233 596 THR-RICH.  
 FT DOMAIN 393 442 ASN-RICH.  
 FT SEQUENCE 1778 AA; 195575 MW; 1D2AED57D927ADC6 CRC64;  
 SQ  
 Query Match 11.5%; Score 88; DB 1; Length 1778;  
 Best Local Similarity 27.3%; Pred. No. 13;  
 Matches 38; Conservative 16; Mismatches 65; Indels 20; Gaps 5;  
 QY- 15 SGSALAGVVPWGGGNGHNGSSGDSSTL-----SIYQGSANAALALQSDARKSETT 69  
 DB 400 SGTWGTGL---FGCANNTANNNTAPTSTFGNNSNFSFGANNNAATKPSFGFGSGTT 456  
 QY 70 ITQSGYNGADVGGADNSTIELTQNGFRNNATYDQLVTRVTHMAHAGGNNAAALVNQT 129  
 DB 457 TTPA--SGPSFGQANNA-----PKPAFGSTATTAPKPACTGLFGGLGAGANTNTATNAT 510  
 QY 130 ASDSSVMVRQVFGNNATA 148  
 DB 511 GTGGSL-----FGNANTA 523  
 RESULT 12  
 ID\_VG38 BPT2 STANDARD; PRT; 262 AA.  
 AC P07875;  
 DT 01-AUG-1988 (Rel. 08, Created)  
 DT 01-AUG-1988 (Rel. 08, Last sequence update)  
 DT 01-AUG-1990 (Rel. 15, Last annotation update)  
 DE Receptor recognizing protein (Protein Gp38).  
 GN 38.  
 OS Bacteriophage T2.  
 OC Viruses; dsDNA viruses, no RNA stage; Caudovirales; Myoviridae;  
 OC T4-like viruses.  
 OX NCBI\_TaxID=10664;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=87283911; PubMed=3302276;  
 RA Riede I., Drexler K., Eschbach M.L., Henning U.;  
 RT "DNA sequence of genes 38 encoding a receptor-recognizing protein of  
 bacteriophages T2, K3 and of K3 host range mutants.";  
 RL J. Mol. Biol. 194:31-39(1987).  
 CC -!- FUNCTION: Vg38 is at the tip of the long tail fibers and serves as  
 the phage recognition site for the cellular receptor.  
 CC -!- MISCELLANEOUS: THIS PHAGE USE OUTER MEMBRANE PROTEINS OMPF AND TTR  
 AS RECEPTORS.  
 CC  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 use by non-profit institutions as long as its content is in no way  
 modified and this statement is not removed. Usage by and for commercial  
 entities requires a license agreement (see <http://www.isb-sib.ch/announce/>)  
 or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch).  
 DR EMBL; X05312; CAA28935.1; -;  
 DR PIR; S00275; S00275.  
 DR InterPro; IPR007932; Tail\_fibre\_GP38.  
 DR Pfam; PF0268; GP38; 1.  
 KW Fiber protein; Phage recognition.  
 SQ SEQUENCE 262 AA; 25801 MW; 056736918F6C745 CRC64;

Query Match 11.3%; Score 86.5; DB 1; Length 262;  
 Best Local Similarity 44.8%; Pred. No. 2;  
 Matches 26; Conservative 4; Mismatches 19; Indels 9; Gaps 3;  
 QY 27 GGGGNGHNGSSGDSSTLTIYQGSANAALALQSDARKSETTITQSGYNGADVGG 84  
 DB 175 GGGGRPFVGGKIGSDSILS-----GSNASL---TDAGTGGTTF-QYAGNGNGVAG 223  
 RESULT 13  
 ID\_GP63 LEIDO STANDARD; PRT; 590 AA.  
 AC P23223;  
 DT 01-NOV-1991 (Rel. 20, Created)  
 DT 10-NOV-1991 (Rel. 20, Last sequence update)  
 DT 10-OCT-2003 (Rel. 42, Last annotation update)  
 DE Leishmanolysin precursor (EC 3.4.24.36) (Cell surface protease)  
 DE (Major surface glycoprotein) (GP63 protein) (Promastigote surface  
 endopeptidase).  
 GN GP63.  
 OS Leishmania donovani.  
 OC Eukaryota; Euglenozoa; Kinetoplastida; Trypanosomatidae; Leishmania.  
 OX NCBI\_TaxID=5661;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX STRAIN=LV9;  
 RX MEDLINE=92107220; PubMed=1762629;  
 RA Webb J.R., Button L.L., McMaster R.W.;  
 RT "Heterogeneity of the genes encoding the major surface glycoprotein  
 of Leishmania donovani.";  
 RL Mol. Biochem. Parasitol. 48:173-184(1991).  
 CC -!- FUNCTION: Has an integral role during the infection of macrophages  
 in the mammalian host.  
 CC -!- CATALYTIC ACTIVITY: Preference for hydrophobic residues at P1 and  
 P1' and basic residues at P2 and P3'. A model nonapeptide is  
 cleaved at -Ala-Tyr-|-Leu-Lys-Lys-  
 CC -!- COPACTOR: Binds 1 zinc ion per subunit (By similarity).  
 CC -!- SUBCELLULAR LOCATION: Attached to the membrane by a GPI-anchor.  
 CC -!- SIMILARITY: Belongs to peptidase family M8.  
 CC  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 use by non-profit institutions as long as its content is in no way  
 modified and this statement is not removed. Usage by and for commercial  
 entities requires a license agreement (see <http://www.isb-sib.ch/announce/>)  
 or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch).  
 DR EMBL; M60048; AAA29244.1; -;  
 DR HSP; P08148; 1LML.  
 DR MEROPS; M08.001; -;  
 DR InterPro; IPR006025; Pept\_M\_Zn\_BS.  
 DR InterPro; IPR001577; Peptidase\_M8.  
 DR Pfam; PF01457; Peptidase\_M8; 1.  
 DR PRINTS; PR00782; LSHMANOLYSIN.  
 DR PROSITE; PS00142; ZINC\_PROTEASE; 1.  
 KW Hydrolase; Metalloprotease; Glycoprotein; Metal-binding; Zinc;  
 Zymogen; Signal; Cell adhesion; GPI-anchor; Lipoprotein.  
 FT SIGNAL 1 39 POTENTIAL  
 FT PROPEP 40 87 ACTIVATION PEPTIDE.  
 FT CHAIN 88 565 LEISHMANOLYSIN.  
 FT PROPEP 566 590 REMOVED IN MATURE FORM (BY SIMILARITY).  
 FT METAL 251 251 ZINC (CATALYTIC) (BY SIMILARITY).  
 FT ACT SITE 252 252 BY SIMILARITY.  
 FT METAL 255 255 ZINC (CATALYTIC) (BY SIMILARITY).  
 FT METAL 321 321 ZINC (CATALYTIC) (BY SIMILARITY).  
 FT DISULFID 112 129 BY SIMILARITY.  
 FT DISULFID 178 217 BY SIMILARITY.  
 FT DISULFID 301 373 BY SIMILARITY.  
 FT DISULFID 380 443 BY SIMILARITY.  
 FT DISULFID 393 412 BY SIMILARITY.  
 FT DISULFID 402 477 BY SIMILARITY.

FT DISULFID 454 498 BY SIMILARITY.  
 FT DISULFID 503 553 BY SIMILARITY.  
 FT DISULFID 523 546 BY SIMILARITY.  
 FT CARBOHYD 287 246 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT LIPID 565 GPI-anchor amidated asparagine (By similarity).  
 SQ SEQUENCE 590 AA; 62950 MW; 0FB315D299659F58 CRC64;  
 Query Match 11.3%; Score 86.5; DB 1; Length 590;  
 Best Local Similarity 47.1%; Pred. No. 4.9;  
 Matches 24; Conservative 4; Mismatches 18; Indels 5; Gaps 2;  
 QY 103 YDQLVTRVTHMAHAGNNAALVNQASDSSVW-VRQVFG-----NNATA 148  
 |||||  
 DB 241 YDQLVTRVTHMAHAGNNAALVNQASDSSVW-VRQVFG-----NNATA 148  
 |||||  
 RESULT 14  
 GP63 LEIME  
 ID GP63 LEIME STANDARD; PRT; 646 AA.  
 AC P43150;  
 DT 01-NOV-1995 (Rel. 32, Created)  
 DT 01-NOV-1995 (Rel. 32, Last sequence update)  
 DT 10-OCT-2003 (Rel. 42, Last annotation update)  
 DE Leishmanolysin C1 precursor (EC 3.4.24.36) (Cell surface protease)  
 DE (Major surface glycoprotein) (GP63 protein) (Promastigote surface  
 endopeptidase).  
 GN GP63-C1.  
 OS Leishmania mexicana.  
 OC Eukaryota; Euglenozoa; Kinetoplastida; Trypanosomatidae; Leishmania.  
 OX NCBI\_TaxID=5665;  
 RN [1]  
 RC SEQUENCE FROM N.A.  
 RC STRAIN=MNYC/BZ/62/M379;  
 RX MEDLINE=93149206; PubMed=8426614;  
 RA Medina-Acosta E., Karsess R.E., Russell D.G.;  
 RT "Structurally distinct genes for the surface protease of Leishmania  
 mexicana are developmentally regulated.";  
 RT Mol. Biochem. Parasitol. 57:31-46(1993).  
 CC -!- FUNCTION: Has an integral role during the infection of macrophages  
 in the mammalian host.  
 CC -!- CATALYTIC ACTIVITY: Preference for hydrophobic residues at P1 and  
 P1' and basic residues at P2 and P3'. A model nonapeptide is  
 cleaved at -Ala-Tyr-I-Leu-Lys-Lys-.  
 CC -!- COFACTOR: Binds 1 zinc ion per subunit (By similarity).  
 CC -!- DEVELOPMENTAL STAGE: Expressed in both the promastigote and the  
 amastigote forms.  
 CC -!- SIMILARITY: Belongs to peptidase family M8.  
 CC  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 the European Bioinformatics Institute. There are no restrictions on its  
 use by non-profit institutions as long as its content is in no way  
 modified and this statement is not removed. Usage by and for commercial  
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC  
 CC EMBL; X64394; CAA45733.1; -;  
 DR PIR; S19916; S19916.  
 DR HSSP; P08148; 1LML.  
 DR MEROPS; M08.001; -;  
 DR GlycoSuiteDB; P43150; -;  
 DR InterPro; IPR006025; Pept. M. Zn BS.  
 DR InterPro; IPR001577; Peptidase M8.  
 DR Pfam; PF01457; Peptidase M8; 1.  
 DR PRINTS; PR00782; LSHMANOLYSIN.  
 DR PROSITE; PS00142; ZINC\_PROTEASE; 1.  
 KW Hydrolase; Metalloprotease; Glycoprotein; Metal-binding; Zinc;  
 KW Zymogen; Signal; Cell adhesion; Multigene family.  
 FT SIGNAL 1 39 POTENTIAL.  
 FT PROPEP 40 102 ACTIVATION PEPTIDE (POTENTIAL).  
 FT CHAIN 103 646 LEISHMANOLYSIN C1.  
 FT METAL 266 286 ZINC (CATALYTIC) (BY SIMILARITY).  
 FT

FT ACT SITE 267 267 BY SIMILARITY.  
 FT METAL 270 270 ZINC (CATALYTIC) (BY SIMILARITY).  
 FT METAL 336 336 ZINC (CATALYTIC) (BY SIMILARITY).  
 FT DISULFID 127 144 BY SIMILARITY.  
 FT DISULFID 193 232 BY SIMILARITY.  
 FT DISULFID 316 388 BY SIMILARITY.  
 FT DISULFID 395 458 BY SIMILARITY.  
 FT DISULFID 408 427 BY SIMILARITY.  
 FT DISULFID 417 492 BY SIMILARITY.  
 FT DISULFID 469 513 BY SIMILARITY.  
 FT DISULFID 518 568 BY SIMILARITY.  
 FT DISULFID 538 561 BY SIMILARITY.  
 FT CARBOHYD 86 86 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 297 297 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 399 399 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 409 409 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 433 433 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 445 445 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 466 466 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 501 501 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 SQ SEQUENCE 646 AA; 69054 MW; FE48DDC78C10B0A CRC64;  
 Query Match 11.2%; Score 86; DB 1; Length 646;  
 Best Local Similarity 94.4%; Pred. No. 6;  
 Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
 QY 103 YDQLVTRVTHMAHAGG 120  
 |||||  
 DB 256 YDQLVTRVTHMAHAGV 273  
 |||||  
 RESULT 15  
 ICEN\_XANCT  
 ID ICEN\_XANCT STANDARD; PRT; 1567 AA.  
 AC P18127;  
 DT 01-NOV-1990 (Rel. 16, Created)  
 DT 01-NOV-1990 (Rel. 16, Last sequence update)  
 DT 16-OCT-2001 (Rel. 40, Last annotation update)  
 DE Ice nucleation protein.  
 GN INAX.  
 OS Xanthomonas campestris (pv. translucens).  
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Xanthomonadales;  
 OC Xanthomonadaceae; Xanthomonas.  
 OX NCBI\_TaxID=343;  
 RN [1]  
 RC SEQUENCE FROM N.A.  
 RC STRAIN=X56S;  
 RX MEDLINE=91080859; PubMed=2259339;  
 RA Zhao J., Orser C.S.;  
 RT "Conserved repetition in the ice nucleation gene inax from  
 Xanthomonas campestris pv. translucens.";  
 RL Mol. Gen. Genet. 223:163-166(1990).  
 CC -!- FUNCTION: Ice nucleation proteins enable bacteria to nucleate  
 crystallization in supercooled water.  
 CC -!- SUBCELLULAR LOCATION: Outer membrane (By similarity).  
 CC -!- DOMAIN: CONTAINS 153 IMPERFECT REPEATS OF THE CONSENSUS  
 OCTAPEPTIDE A-G-Y-G-S-T-L-T; FURTHER ON A 16-RESIDUE AND A  
 REGIONAL 48-RESIDUE PERIODICITY IS SUPERIMPOSED.  
 CC -!- MISCELLANEOUS: A STRUCTURAL MODEL IS SUGGESTED IN WHICH THE ICE  
 NUCLEATION PROTEIN DISPLAYS A SYMMETRY RELATED TO THAT OF ICE.  
 CC -!- SIMILARITY: Belongs to the bacterial ice nucleation protein  
 family.  
 CC  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 the European Bioinformatics Institute. There are no restrictions on its  
 use by non-profit institutions as long as its content is in no way  
 modified and this statement is not removed. Usage by and for commercial  
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC  
 CC EMBL; X52970; CAA37140.1; -;  
 DR HSSP; P06620; 1INA.  
 DR

```
DR InterPro; IPR000258; Ice.nucleatn.
DR Pfam; PF00818; Ice.nucleation; 81.
DR PRINTS; PR00327; ICENUCLEATN.
DR PROSITE; PS00314; ICE NUCLEATION; 57.
KW Ice nucleation; Repeat; Outer membrane.
SQ SEQUENCE 1567 AA; 152548 MW; C8B451D959ECAD63 CRC64;

Query Match      11.1%; Score 85.5; DB 1; Length 1567;
Best Local Similarity 26.4%; Pred. No. 18;
Matches 39; Conservative 27; Mismatches 39; Indels 43; Gaps 10;

QY 34 GGNSSGPDSTLSIYQYGSANAALALQSDARKSETTITQSGYCN-----GADV----- 81
Db 1055 GSTGTAGADSTL-IAGYGSTQTA-----GSDSSLT-AGYGTQTARQSGSDITAGYGS 1104

QY 82 --CGGADNSTIE---LTQN-GFRNNATYDQLVTRVVTHEMA-----HAGNNRAALV- 126
Db 1105 TGTAGADSSLIAGYGSTQTAGYDSNLTAGYGTQTAREDSLSLTAGYGTSTAGHDSLLIA 1164

QY 127 ----NOTASDSSVMVRQVGFNNATANQ 150
Db 1165 GYGSTQTAGYNSILT--TGYGSTQTAE 1190
```

Search completed: August 2, 2004, 14:49:33  
Job time : 6.3 secs



GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: August 2, 2004, 14:39:17 ; Search time 29.7 Seconds  
(without alignments)  
1604.150 Million cell updates/sec

Title: US-09-543-407-30  
Perfect score: 768  
Sequence: 1 MKLLKVAFAAIWSSGALA.....DSSVMVRQVGFNNATANYQ 151

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 1017041 seqs, 315518202 residues

Total number of hits satisfying chosen parameters: 1017041

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000  
Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : SPTREMBL\_25:\*  
1: sp\_archaea:\*  
2: sp\_bacteria:\*  
3: sp\_fungi:\*  
4: sp\_human:\*  
5: sp\_invertebrate:\*  
6: sp\_mammal:\*  
7: sp\_mhc:\*  
8: sp\_organelle:\*  
9: sp\_phase:\*  
10: sp\_plant:\*  
11: sp\_rodent:\*  
12: sp\_virus:\*  
13: sp\_vertebrate:\*  
14: sp\_unclassified:\*  
15: sp\_rvrius:\*  
16: sp\_bacteriap:\*  
17: sp\_cheap:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	ID	Description
1	672	87.5	152	2	O33802 salmonella
2	570.5	74.3	150	2	Q7X243 citrobacter
3	534	69.5	149	2	Q7X240 citrobacter
4	506.5	66.0	152	16	Q8CW63 escherichia
5	435.5	56.7	150	2	Q7X237 enterobacte
6	380	49.5	76	2	Q54069 salmonella
7	126	16.4	502	16	Q8EIH4 shewanella
8	122	15.9	29	2	Q8S3J5 escherichia
9	112	14.6	139	16	Q8EIH3 shewanella
10	111.5	14.5	171	16	Q89J13 bradyrhizob
11	105	13.7	130	16	Q89J14 bradyrhizob
12	104.5	13.6	151	16	Q7UCZ1 shigella fl
13	104.5	13.6	160	16	Q8CW64 escherichia
14	104.5	13.5	160	16	Q83U7 shigella fl
15	104	13.5	157	16	Q88H70 pseudomonas
16	102.5	13.3	3659	16	Q98LH6 rhizobium l

17	102	13.3	1422	16	Q8EFU3
18	101.5	13.2	531	13	Q8AWA8
19	101	13.2	1748	5	Q94821
20	101	13.2	7716	16	Q7UWZ8
21	99.5	13.0	1209	16	Q89CK5
22	99.5	13.0	2174	16	Q92U08
23	99	12.9	362	16	Q89D03
24	99	12.9	441	2	O87327
25	98.5	12.8	348	13	Q93397
26	98.5	12.8	624	3	Q8NIV1
27	98	12.8	145	16	Q8UGN9
28	97.5	12.7	91	2	Q8S3J8
29	97.5	12.7	477	2	Q7X485
30	97.5	12.7	586	16	Q8EXJ2
31	97.5	12.7	645	16	Q7UIC5
32	97.5	12.7	646	16	O53818
33	97	12.6	1410	16	Q8CMJ0
34	96.5	12.6	191	10	Q7XDR3
35	96.5	12.6	313	10	Q9SYZ5
36	95.5	12.4	151	2	Q7X244
37	95.5	12.4	154	16	Q89J15
38	95.5	12.4	346	5	Q9SRS5
39	95.5	12.4	908	5	Q9VM71
40	95	12.4	151	2	Q7X238
41	95	12.4	196	10	O22638
42	94.5	12.3	153	16	Q89J16
43	94.5	12.3	908	5	Q9VBC4
44	94.5	12.3	909	5	Q8T4E0
45	94.5	12.3	1713	3	Q8TGE1

ALIGNMENTS

RESULT 1  
O33802 PRELIMINARY; PRT; 152 AA.  
ID O33802  
AC O33802  
DT 01-JAN-1998 (TRENBLrel. 05, Created)  
DT 01-JAN-1998 (TRENBLrel. 05, Last sequence update)  
DT 01-DEC-2001 (TRENBLrel. 19, Last annotation update)  
DE AgfA protein (Fragment).  
GN AGFA.  
OS Salmonella typhimurium.  
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;  
OC Enterobacteriaceae; Salmonella.  
OX NCBI\_TaxID=602;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=98053981; PubMed=9393832;  
RA Sukupolvi S.S., Lorentz R.G., Gordon J.I., Bian Z., Pfeifer J.D.,  
RA Normark S.J., Rhen M.;  
RT "Expression of thin, aggregative fimbriae promotes interaction of  
RT Salmonella typhimurium SR-11 with mouse small intestinal epithelial  
RT cells.";  
RL Infect. Immun. 65:5320-5325(1997).  
DR EMBL; AJ000514; CAA04151.1; -  
FT NON TER 152  
SQ SEQUENCE 152 AA; 15401 MW; 9DA7DADC2364B006 CRC64;

QY	1	MKLLKVAFAAIWSSGALA	GVVPQWGGGNGHNGSGSPDSTLSIYQGSANAALAIQ	60
DB	1	MKLLKVAFAAIWSSGSA	VGVVPQWGGGNGHNGSGSPDSTLSIYQGSANAALAIQ	60
QY	61	SDARKSETTITQSGYGN	GADVCGQADNSTIELTQNGFNATIDQWNAKNSDITVGYGG	120
DB	61	SDARKSETTITQSGYGN	GADVCGQADNSTIELTQNGFNATIDQWNAKNSDITVGYGG	120
QY	121	NNAALVNQATSDSSVM	VRQVGFNNATANYQ	151

Query Match 87.5%; Score 672; DB 2; Length 152;  
Best Local Similarity 89.4%; Pred. No. 3e-46;  
Matches 135; Conservative 3; Mismatches 13; Indels 0; Gaps 0;

Db 121 NNAALVNOTASDSSVMVRQVGFNNAPANQY 151  
 |||||

## RESULT 2

Q7X243 Q7X243 PRELIMINARY; PRT; 150 AA.  
 AC Q7X243;  
 DT 01-OCT-2003 (TReMBLrel. 25, Created)  
 DT 01-OCT-2003 (TReMBLrel. 25, Last sequence update)  
 DT 01-OCT-2003 (TReMBLrel. 25, Last annotation update)  
 DE Curlin-csgA protein.  
 GN CSGA.  
 OS Citrobacter sp. Fec2.  
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;  
 OC Enterobacteriaceae; Citrobacter.  
 OX NCBI\_TaxID=213763;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=Fec2;  
 RA Zogaj X., Bokranz W., Nimitz M., Romling U.;  
 RT "Production of Cellulose and Curlin Fimbriae by Members of the Family  
 Enterobacteriaceae Isolated from the Human Gastrointestinal Tract.";  
 RL Infect. Immun. 72:4151-4158 (2003).  
 DR EMBL; AJ515700; CAD56672.1; -  
 SQ SEQUENCE 150 AA; 15016 MW; 1D7141B8D6973DC6 CRC64;

Query Match 74.3%; Score 570.5; DB 2; Length 150;  
 Best Local Similarity 77.5%; Pred. No. 3.7e-38;  
 Matches 117; Conservative 14; Mismatches 19; Indels 1; Gaps 1;  
 QY 1 MLLKVAAPAAIVVSGSALAGVVPQWGGGNNHGGGSGSPDSTLSIYQYGSANALALQ 60  
 DB 1 MLLQVAAPAAIVVSGSALAGVVPQWGGG -GGGSSGSPSTLSIYQYGVNNAALALQ 59  
 QY 61 SDARKSETTITQSGYNGADVGQADNSTLTQTQNGFRNNATYDQLVTRVVTHEMAHAG 120  
 DB 60 SDARKSDTTIHQNGFGNGADVGQSDNSTLDTQTQNGFKNNATIDQWNGKNSDITVSQYGG 119

Query Match 74.3%; Score 570.5; DB 2; Length 150;  
 Best Local Similarity 77.5%; Pred. No. 3.7e-38;  
 Matches 117; Conservative 14; Mismatches 19; Indels 1; Gaps 1;  
 QY 121 NNAALVNOTASDSSVMVRQVGFNNATANQY 151  
 DB 120 HNAALVNOTASDSSVLRVHQVGFNNATANQY 150

## RESULT 3

Q7X240 Q7X240 PRELIMINARY; PRT; 149 AA.  
 AC Q7X240;  
 DT 01-OCT-2003 (TReMBLrel. 25, Created)  
 DT 01-OCT-2003 (TReMBLrel. 25, Last sequence update)  
 DT 01-OCT-2003 (TReMBLrel. 25, Last annotation update)  
 DE Curlin-csgA protein.  
 GN CSGA.  
 OS Citrobacter freundii.  
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;  
 OC Enterobacteriaceae; Citrobacter.  
 OX NCBI\_TaxID=546;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=Fec4;  
 RA Zogaj X., Bokranz W., Nimitz M., Romling U.;  
 RT "Production of Cellulose and Curlin Fimbriae by Members of the Family  
 Enterobacteriaceae Isolated from the Human Gastrointestinal Tract.";  
 RL Infect. Immun. 72:4151-4158 (2003).  
 DR EMBL; AJ515701; CAD56675.1; -  
 SQ SEQUENCE 149 AA; 15260 MW; 946DD52017F648FD CRC64;

Query Match 69.5%; Score 534; DB 2; Length 149;  
 Best Local Similarity 72.2%; Pred. No. 3e-35;  
 Matches 109; Conservative 16; Mismatches 24; Indels 1; Gaps 1;  
 QY 1 MLLKVAAPAAIVVSGSALAGVVPQWGGGNNHGGGSGSPDSTLSIYQYGSANALALQ 60  
 |||||

Db 1 MLLKVAAPAAIVVSGSALAGVVPQW--GGNHGGGSGYGPDSLSIYQYGSNNANALQ 58  
 QY 61 SDARKSETTITQSGYNGADVGQADNSTLTQTQNGFRNNATYDQLVTRVVTHEMAHAG 120  
 DB 59 SDARKSDVTITQHGNGGAVVVGQADDSTISLQTFQNSATIDQWNAKNAKADISVTQFGG 118  
 QY 121 NNAALVNOTASDSSVMVRQVGFNNATANQY 151  
 DB 119 RNgALVNOTASDSSVLRVHQVGFNNATANQY 149

## RESULT 4

Q8CW63 Q8CW63 PRELIMINARY; PRT; 152 AA.  
 AC Q8CW63;  
 DT 01-MAR-2003 (TReMBLrel. 23, Created)  
 DT 01-MAR-2003 (TReMBLrel. 23, Last sequence update)  
 DT 01-MAR-2003 (TReMBLrel. 23, Last annotation update)  
 DE Major curlin subunit precursor.  
 GN CSGA OR C1306.  
 OS Escherichia coli O6.  
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;  
 OC Enterobacteriaceae; Escherichia.  
 OX NCBI\_TaxID=217992;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=O6:H1 / CFT073 / ATCC 700928;  
 EX MEDLINE=22388234; PubMed=12471157;  
 RA Welch R.A., Burland V., Plunkett G. III, Redford P., Roesch P.,  
 RA Raako D., Buckles E.J., Liou S.-R., Boutin A., Hackett J., Stroud D.,  
 RA Mayhew G.F., Rose D.J., Zhou S., Schwartz D.C., Perna N.T.,  
 RA Mobley H.L.T., Donnenberg M.S., Blattner F.R.;  
 RT "Extensive mosaic structure revealed by the complete genome sequence  
 of uropathogenic Escherichia coli.";  
 RL Proc. Natl. Acad. Sci. U.S.A. 99:17020-17024 (2002).  
 DR EMBL; AE016759; AAN79779.1; -  
 KW Complete proteome.  
 SQ SEQUENCE 152 AA; 15064 MW; 3BA57F34C1240B83 CRC64;

Query Match 66.0%; Score 506.5; DB 16; Length 152;  
 Best Local Similarity 69.1%; Pred. No. 4.9e-33;  
 Matches 105; Conservative 17; Mismatches 29; Indels 1; Gaps 1;  
 QY 1 MLLKVAAPAAIVVSGSALAGVVPQW-GGGNNHGGGSGSPDSTLSIYQYGSANALAL 59  
 DB 1 MLLKVAAPAAIVVSGSALAGVVPQYGGGNGHGGGSGNSPNSLNIYQYGGNSALAQ 60

QY 60 QSDARKSETTITQSGYNGADVGQADNSTLTQTQNGFRNNATYDQLVTRVVTHEMAHAG 119  
 DB 61 QADARNSDLTITQHGNGGADVGQSDSDSDLTQTQNGFSATLDQWNGKDSITWVKQFG 120  
 QY 120 GNNALVNOTASDSSVMVRQVGFNNATANQY 151  
 DB 121 GGNGAANDQTAASNSVNVTVQVGFNNATAHQY 152

## RESULT 5

Q7X237 Q7X237 PRELIMINARY; PRT; 150 AA.  
 AC Q7X237;  
 DT 01-OCT-2003 (TReMBLrel. 25, Created)  
 DT 01-OCT-2003 (TReMBLrel. 25, Last sequence update)  
 DT 01-OCT-2003 (TReMBLrel. 25, Last annotation update)  
 DE Curlin-csgA protein.  
 GN CSGA.  
 OS Enterobacter sakazakii.  
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;  
 OC Enterobacteriaceae; Enterobacter.  
 OX NCBI\_TaxID=28141;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=Fec39;  
 RA Zogaj X., Bokranz W., Nimitz M., Romling U.;

RT "Production of Cellulose and Curli Fimbriae by Members of the Family  
 RL Enterobacteriaceae Isolated from the Human Gastrointestinal Tract."  
 RL Infect. Immun. 72:4151-4158(2003).  
 DR EMBL: AJ515702; CAD56678.1; -.  
 SQ SEQUENCE 150 AA; 15112 MW; 5DBBB2D872DF15F3 CRC64;

Query Match 56.7%; Score 435.5; DB 2; Length 150;  
 Best Local Similarity 60.3%; Pred. No. 2.2e-27;  
 Matches 91; Conservative 26; Mismatches 33; Indels 1; Gaps 1;

QY 1 MKLLKVAAPAAIIVSGSALAGVVPWGGGNGHNGGNSGPDSTLSIYQYGSANAALAIQ 60  
 Db 1 MAFIKVAALAIIVSGSAGWAGMINO-GGWCHGHGGYGGFNPSTLNIYQGGNSALAIQ 59  
 QY 61 SPARKSETTITQSGYNGADVGCGADNSTIETQNGFRNNATYDQIVTVRVVTHEMAHAGG 120  
 Db 60 TDARNSVLNIQTGGGNGADVGCGSDSSINLTQNGFGNSATLDQNSKDSVMNVSYQGG 119  
 QY 121 NNAALVNOTASDSSVMVRQVGGNNATANQY 151  
 Db 120 LAGALVDQTASNSTVNVVTQIGFNGHATAHQY 150

RESULT 6  
 Q54069 ID Q54069 PRELIMINARY; PRT; 76 AA.  
 AC Q54069;  
 DT 01-NOV-1996 (TREMBlrel. 01, Created)  
 DT 01-NOV-1996 (TREMBlrel. 01, Last sequence update)  
 DT 01-DEC-2001 (TREMBlrel. 19, Last annotation update)  
 DE SEF17 fimbria (Fragment).  
 GN AGFA.  
 OS Salmonella enteritidis.  
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;  
 OC Enterobacteriaceae; Salmonella.  
 OX NCBI\_TaxID=592;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=SE30;  
 RA Cox J.M., Eglesos S., Woolcock J.B.;  
 RT "Virulence of Salmonella enteritidis in chickens correlates with  
 RT colony morphology and expression of SEF17 fimbriae."  
 RL Submitted (APR-1996) to the EMBL/GenBank/DBJ databases.  
 DR EMBL: U53207; AAA98671.1; -.  
 FT NON\_TER 1 1  
 FT NON\_TER 76 76  
 SQ SEQUENCE 76 AA; 7704 MW; 2FD5411241A7BCB1 CRC64;

Query Match 49.5%; Score 380; DB 2; Length 76;  
 Best Local Similarity 97.4%; Pred. No. 2.7e-23;  
 Matches 74; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 30 GNHNGGNSGPDSTLSIYQYGSANAALAIQSDARKSETTITQSGYNGADVGCGADNST 89  
 Db 1 GNHNGGNSGPDSTLSIYQYGSANAALAIQSDARKSETTITQSGYNGADVGCGADNST 60  
 QY 90 IELTQNGFRNNATYDQ 105  
 Db 61 IELTQNGFRNNATIDQ 76

RESULT 7  
 Q8EIH4 ID Q8EIH4 PRELIMINARY; PRT; 502 AA.  
 AC Q8EIH4;  
 DT 01-MAR-2003 (TREMBlrel. 23, Created)  
 DT 01-MAR-2003 (TREMBlrel. 23, Last sequence update)  
 DT 01-MAR-2003 (TREMBlrel. 23, Last annotation update)  
 DE Conserved hypothetical protein.  
 GN SO0865.  
 OS Shewanella oneidensis.  
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Alteromonadales;  
 OC Alteromonadaceae; Shewanella.

OX NCBI\_TaxID=70863;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=MR-1;  
 RX MEDLINE=22297686; PubMed=12368813;  
 RA Heidelberg J.F., Paulsen I.T., Nelson K.E., Gaidos E.J., Nelson W.C.,  
 RA Read T.D., Eisen J.A., Seshadri R., Wad N., Methe B., Clayton R.A.,  
 RA Meyer T., Tsapin A., Scott J., Beanan M., Brinkac L., Daugherty S.,  
 RA DeBoy R.T., Dodson R.J., Durkin A.S., Haft D.H., Kolonay J.F.,  
 RA Madupu R., Peterson J.D., Umayam L.A., White O., Wolf A.M.,  
 RA Vamathevan J., Weidman J., Impraim M., Lee K., Berry K., Lee C.,  
 RA Mueller J., Khouri H., Gill J., Utterback T.R., McDonald L.A.,  
 RA Feldblyum T.V., Smith H.O., Venter J.C., Nealon K.H., Fraser C.M.;  
 RT "Genome sequence of the dissimilatory metal ion-reducing bacterium  
 RT Shewanella oneidensis";  
 RL Nat. Biotechnol. 20:1118-1123(2002).  
 DR EMBL: AB015532; AAN53941.1; -.  
 DR TIGR: SO0865; -.  
 KW Hypothetical protein; Complete proteome.  
 SQ SEQUENCE 502 AA; 52441 MW; D08CA23D6C46B62D CRC64;

Query Match 15.4%; Score 126; DB 16; Length 502;  
 Best Local Similarity 27.0%; Pred. No. 0.048;  
 Matches 41; Conservative 22; Mismatches 57; Indels 32; Gaps 5;

QY 29 GGNHNG-----GCG-----SSGPDSTLSIYQYGSANA---ALALQS 61  
 Db 231 GDNHTGFVYALAGSENDISMEQEGSNNTAYLSMTTGDNTVDITQDGSNTVGDSLIADI 290  
 QY 62 DARKSETTITQSGYNGADVGCGADNSTIETQNGFRNNATYDQIVTVRVVTHEMAHAGG 121  
 Db 291 QGDDNDITIKQKGSNGAEFQVWGDSNDVDLKQGDANFATFGAYGTD-NDPDLSSKGDN 349

QY 122 NAAALVNOTASDSSVMVRQVGGN-----NATAN 149  
 Db 350 NELVAFATGECDSIEISQEGDANFAYVDATCN 381

RESULT 8  
 Q9S3J5 ID Q9S3J5 PRELIMINARY; PRT; 29 AA.  
 AC Q9S3J5;  
 DT 01-MAY-2000 (TREMBlrel. 13, Created)  
 DT 01-MAY-2000 (TREMBlrel. 13, Last sequence update)  
 DT 01-MAR-2001 (TREMBlrel. 16, Last annotation update)  
 DE Curlin subunit monomer (Fragment).  
 GN CSGA.  
 OS Escherichia coli.  
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;  
 OC Enterobacteriaceae; Escherichia.  
 OX NCBI\_TaxID=562;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TRANSPOSON=Insertion sequence IS1;  
 RX MEDLINE=99314153; PubMed=10386375;  
 RA La Ragione R.M., Collighan R.J., Woodward M.J.;  
 RT "Non-curlation of Escherichia coli O78:K80 isolates associated with  
 RT IS1 insertion on in csgr and reduced persistence in poultry infection."  
 RL FEMS Microbiol. Lett. 175:247-253(1999).  
 DR EMBL: AJ131756; CAB45380.1; -.  
 FT NON\_TER 29 29  
 SQ SEQUENCE 29 AA; 2789 MW; E290DFC07ABBB243 CRC64;

Query Match 15.9%; Score 122; DB 2; Length 29;  
 Best Local Similarity 89.7%; Pred. No. 0.0035;  
 Matches 26; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 MKLLKVAAPAAIIVSGSALAGVVPWGGG 29  
 Db 1 MKLLKVAATAIIVFSGSALAGVVPYGGG 29

RESULT 9

Q8EIH3 PRELIMINARY; PRT; 139 AA.

AC Q8EIH3; 01-MAR-2003 (TrEMBLrel. 23, Created)

DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)

DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)

DE Minor curlin subunit CsgB, putative.

GN S00866.

OS Shewanella oneidensis.

OC Bacteria; Proteobacteria; Gammaproteobacteria; Alteromonadales;

OC Alteromonadaceae; Shewanella.

OX NCBI\_TaxID=70863;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=MR-1;

RX MEDLINE=2297686; PubMed=12368813;

RA Heidelberg J.F., Paulsen I.T., Nelson K.E., Gaidos E.J., Nelson W.C.,

RA Read T.D., Eissen J.A., Seshadri R., Ward N., Methe B., Clayton R.A.,

RA Meyer T., Tsapin A., Scott J., Beanan M., Brinkac L., Daugherty S.,

RA DeBoy R.T., Dodson R.J., Durkin A.S., Haft D.H., Kolonay J.F.,

RA Madupu R., Peterson J.D., Umayam L.A., White O., Wolf A.M.,

RA Vamathevan J., Weidman J., Imprim M., Lee K., Berry K., Lee C.,

RA Mueller J., Khouri H., Gill J., Utterback T.R., McDonald L.A.,

RA Feldblyum T.V., Smith H.O., Venter J.C., Nealeson K.H., Fraser C.M.,

RA "Genome sequence of the dissimilatory metal ion-reducing bacterium

RT Shewanella oneidensis";

RL Nat. Biotechnol. 20:1118-1123 (2002).

DR EMBL; AE015532; AAN53942.1; -

DW TIGR; S00866; -

KW Complete proteome.

SQ SEQUENCE 139 AA; 14811 MW; 41EC1CPA76957920 CRC64;

Query Match 14.6%; Score 112; DB 16; Length 139;

Best Local Similarity 28.3%; Pred. No. 0.14;

Matches 32; Conservative 22; Mismatches 45; Indels 14; Gaps 3;

QY 39 SGPDSLISYQVGSANAALQSDARKSETTITQSGYNGADVQGGADNSTELTQNGPR 98

DB 41 SGRNLDLVQQTANQGVIFQSGSNS-AVITQAGNDNISLVITQIGTNEVQLQVGAQ 99

QY 99 NNATYQDLVTRVVTTHMAHAGGNAALVNQTSDDSSVMYRVQVFGNNATANQY 151

DB 100 NKASITQI-----GNDNLVQLNQLGS-GNFSIQIQAAGAAISITQY 139

RESULT 10

Q89JI3 PRELIMINARY; PRT; 171 AA.

AC Q89JI3; 01-JUN-2003 (TrEMBLrel. 24, Created)

DT 01-JUN-2003 (TrEMBLrel. 24, Last sequence update)

DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)

DE CsgA protein.

GN CSGA OR BLS300.

OS Bradyrhizobium japonicum.

OC Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;

OC Bradyrhizobiaceae; Bradyrhizobium.

OX NCBI\_TaxID=375;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=USDA 110;

RX MEDLINE=22484998; PubMed=12597275;

RA Kaneko T., Nakamura Y., Sato S., Minamisawa K., Uchiumi T.,

RA Sasamoto S., Watanabe A., Idesawa K., Iriguchi M., Kawashima K.,

RA Kohara M., Matsumoto M., Shimo S., Tsuruoka H., Wada T., Yamada M.,

RA Tabata S.;

RA "Complete genomic sequence of nitrogen-fixing symbiotic bacterium

RT Bradyrhizobium japonicum USDA110.";

RL DNA Res. 9:189-197 (2002).

DR EMBL; AP005954; BAC50564.1; -

DW Complete proteome.

KW Complete proteome.

SQ SEQUENCE 171 AA; 17448 MW; 995DB08C01498381 CRC64;

Query Match 13.7%; Score 105; DB 16; Length 130;

Best Local Similarity 24.7%; Pred. No. 0.46;

Matches 37; Conservative 27; Mismatches 60; Indels 26; Gaps 4;

QY 4 LKVAFAAIVVSGSALAGVVPQGGGNGHNGGNSGPDSTLSIYQVGSANAALQSDA 63

DB 1 MRITYLVATAIALSALTIVDAQ-----AGNSA-----SVLQFGTNSFSIQTGS 45

QY 64 RKSETTITQSGYNGADVQGGADNSTIELT---QNGPRNATYDQLVTRVVTTHMAHAGG 120

DB 46 TSNNATTLQFGATNTATTLQTSLLTNTAVTQGGTATASNTALTQGV-----GG 97

QY 121 NNAALVNQTSDDSSVMYRVQVFGNNATANQ 150

DB 98 SNSSLIGQIGNNTAGVQGLINGSTILO 127

RESULT 12

Q7UCZ1 PRELIMINARY; PRT; 151 AA.

AC Q7UCZ1; 01-OCT-2003 (TrEMBLrel. 25, Created)

DT 01-OCT-2003 (TrEMBLrel. 25, Last sequence update)

DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)

DE Minor curlin subunit.

GN CSGB OR S1108.

OS Shigella flexneri.

OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;

Query Match 14.5%; Score 111.5; DB 16; Length 171;

Best Local Similarity 24.8%; Pred. No. 0.19;

Matches 40; Conservative 24; Mismatches 54; Indels 43; Gaps 5;

QY 1 MKLLKVAFAAIVVSGSALAGVVPQGGGNGHNGGNSGPDSTLSIYQVGSANAALQ 60

DB 40 MRKLLFFASVAVLALSSAAQA-----NTSTTVQVGLVNGSVTQ 78

QY 61 SDARKSETTITQSGYNGADVQGGAD---NSTIELTQNGPRNATYDQLVTRVVTTHMA 116

DB 79 NGLTNDSSITQIGLNGASTMQGTSPLNNVSTVQAGVQNSATTGQV-----128

QY 117 HAGGNAALVNQTA-----SDSSVMYRVQVFG-NNATANQ 150

DB 129 -AFGNGNSAITQNGFGPPALQNGNSASVQLSFGINTSTVSQ 168

RESULT 11

Q89JI4 PRELIMINARY; PRT; 130 AA.

AC Q89JI4; 01-JUN-2003 (TrEMBLrel. 24, Created)

DT 01-JUN-2003 (TrEMBLrel. 24, Last sequence update)

DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)

DE BLS5299 protein.

GN BLS5299.

OS Bradyrhizobium japonicum.

OC Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;

OC Bradyrhizobiaceae; Bradyrhizobium.

OX NCBI\_TaxID=375;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=USDA 110;

RX MEDLINE=22484998; PubMed=12597275;

RA Kaneko T., Nakamura Y., Sato S., Minamisawa K., Uchiumi T.,

RA Sasamoto S., Watanabe A., Idesawa K., Iriguchi M., Kawashima K.,

RA Kohara M., Matsumoto M., Shimo S., Tsuruoka H., Wada T., Yamada M.,

RA Tabata S.;

RA "Complete genomic sequence of nitrogen-fixing symbiotic bacterium

RT Bradyrhizobium japonicum USDA110.";

RL DNA Res. 9:189-197 (2002).

DR EMBL; AP005954; BAC50564.1; -

DW Complete proteome.

KW Complete proteome.

SQ SEQUENCE 130 AA; 12699 MW; ACFB2D66A48D260F CRC64;

Query Match 13.7%; Score 105; DB 16; Length 130;

Best Local Similarity 24.7%; Pred. No. 0.46;

Matches 37; Conservative 27; Mismatches 60; Indels 26; Gaps 4;

QY 4 LKVAFAAIVVSGSALAGVVPQGGGNGHNGGNSGPDSTLSIYQVGSANAALQSDA 63

DB 1 MRITYLVATAIALSALTIVDAQ-----AGNSA-----SVLQFGTNSFSIQTGS 45

QY 64 RKSETTITQSGYNGADVQGGADNSTIELT---QNGPRNATYDQLVTRVVTTHMAHAGG 120

DB 46 TSNNATTLQFGATNTATTLQTSLLTNTAVTQGGTATASNTALTQGV-----GG 97

QY 121 NNAALVNQTSDDSSVMYRVQVFGNNATANQ 150

DB 98 SNSSLIGQIGNNTAGVQGLINGSTILO 127

RESULT 12

Q7UCZ1 PRELIMINARY; PRT; 151 AA.

AC Q7UCZ1; 01-OCT-2003 (TrEMBLrel. 25, Created)

DT 01-OCT-2003 (TrEMBLrel. 25, Last sequence update)

DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)

DE Minor curlin subunit.

GN CSGB OR S1108.

OS Shigella flexneri.

OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;



```

OC Enterobacteriaceae; Shigella.
OX NCBI_TaxID=623;
RN [1]_
RP SEQUENCE FROM N.A.
RC STRAIN=2457T / ATCC 700930 / Serotype 2a;
RX MEDLINE=22590274; PubMed=12704152; Venkatesan M.M., Deng W.,
RA Wei J., Goldberg M.B., Burland V.,
RA Fournier G., Mayhew G.F., Plunkett G. III, Rose D.J., Darling A.,
RA Mau B., Perna N.T., Payne S.M., Runyen-Janecky L.J., Zhou S.,
RA Schwartz D.C., Blattner F.R.;
RT "Complete genome sequence and comparative genomics of Shigella
RL flexneri serotype 2a strain 2457T.";
DR EMBL; AB016981; AAP16542.1; -;
SQ SEQUENCE 151 AA; 15868 MW; 5D5D266B964014A0 CRC64;

Query Match 13.6%; Score 104.5; DB 16; Length 151;
Best Local Similarity 29.1%; Pred. No. 0.65;
Matches 34; Conservative 15; Mismatches 57; Indels 11; Gaps 3;

QY 38 SSGPDSTLSIYQGSANAALALQSDARKSETTITQSGYNGADVGGADNSTIELTQNGF 97
Db 21 AAGYDLANSEYNF----AVNELSKSFNQAAIIGAGTNNSAQLRQGGSKLLAVVAQEGS 76
QY 98 RNNATYDQLVTRVVTHEMAH---AGGNNALVNQTASDSSVMVRQVGFNNATANQY 151
Db 77 SNRAKIDQ-----TGDYNLAYIDQAGSANDASISQGYGNTAMIIQKSGNKANITQY 129

RESULT 13
O8CM64
ID Q8CW64 PRELIMINARY; PRT; 160 AA.
AC Q8CW64;
DT 01-MAR-2003 (TrEMBLrel. 23, Created)
DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE Minor curlin subunit precursor.
GN CSGB OR C1305
OS Escherichia coli O6.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Escherichia.
OX NCBI_TaxID=217992;
RN [1]_
RP SEQUENCE FROM N.A.
RC STRAIN=O6:H1 / CFT073 / ATCC 700928;
RX MEDLINE=2238234; PubMed=12471157;
RA Welch R.A., Burland V., Plunkett G. III, Redford P., Roesch P.,
RA Rasko D., Buckles E.L., Liou S.-R., Boutin A., Hackett J., Stroud D.,
RA Mayhew G.F., Rose D.J., Zhou S., Schwartz D.C., Perna N.T.,
RA Mobley H.L.T., Donnenberg M.S., Blattner F.R.;
RT "Extensive mosaic structure revealed by the complete genome sequence
RT of uropathogenic Escherichia coli.";
RL Proc. Natl. Acad. Sci. U.S.A. 99:17020-17024 (2002).
DR EMBL; AB016759; AAN79778.1; -;
KW Complete proteome.
SQ SEQUENCE 160 AA; 16963 MW; 49F68448D979B986 CRC64;

Query Match 13.6%; Score 104.5; DB 16; Length 160;
Best Local Similarity 29.1%; Pred. No. 0.65;
Matches 34; Conservative 15; Mismatches 57; Indels 11; Gaps 3;

QY 38 SSGPDSTLSIYQGSANAALALQSDARKSETTITQSGYNGADVGGADNSTIELTQNGF 97
Db 30 AAGYDLANSEYNF----AVNELSKSFNQAAIIGAGTNNSAQLRQGGSKLLTVAQEGS 85
QY 98 RNNATYDQLVTRVVTHEMAH---AGGNNALVNQTASDSSVMVRQVGFNNATANQY 151
Db 86 SNRAKIDQ-----TGDYNLAYIDQAGSANDASISQGYGNTAMIIQKSGNKANITQY 138

RESULT 14
O83RUT
ID Q83RUT PRELIMINARY; PRT; 160 AA.

```

```

Q83RUT;
AC 01-JUN-2003 (TrEMBLrel. 24, Created)
DT 01-JUN-2003 (TrEMBLrel. 24, Last sequence update)
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE Minor curlin subunit precursor, similar to CsgA.
GN CSGB OR SF1035.
OS Shigella flexneri.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Shigella.
OX NCBI_TaxID=623;
RN [1]_
RP SEQUENCE FROM N.A.
RC STRAIN=301 / Serotype 2a;
RX MEDLINE=22272406; PubMed=12384590;
RA Jin Q., Yuan Z., Xu J., Wang Y., Shen Y., Lu W., Wang J., Liu H.,
RA Yang J., Yang F., Zhang X., Zhang J., Yang G., Wu H., Ou D., Dong J.,
RA Sun L., Xue Y., Zhao A., Zhu J., Kan B., Ding K., Chen S.,
RA Cheng H., Yao Z., He B., Chen R., Ma D., Qiang B., Wen Y., Hou Y.,
RA Yu J.;
RT "Genome sequence of Shigella flexneri 2a: insights into pathogenicity
RT through comparison with genomes of Escherichia coli K12 and O157.";
RL Nucleic Acids Res. 30:4432-4441 (2002).
DR EMBL; AB015131; AAM42658.1; -;
KW Complete proteome.
SQ SEQUENCE 160 AA; 16919 MW; 50269F5268D2A32F CRC64;

Query Match 13.6%; Score 104.5; DB 16; Length 160;
Best Local Similarity 29.1%; Pred. No. 0.65;
Matches 34; Conservative 15; Mismatches 57; Indels 11; Gaps 3;

QY 38 SSGPDSTLSIYQGSANAALALQSDARKSETTITQSGYNGADVGGADNSTIELTQNGF 97
Db 30 AAGYDLANSEYNF----AVNELSKSFNQAAIIGAGTNNSAQLRQGGSKLLAVVAQEGS 85
QY 98 RNNATYDQLVTRVVTHEMAH---AGGNNALVNQTASDSSVMVRQVGFNNATANQY 151
Db 86 SNRAKIDQ-----TGDYNLAYIDQAGSANDASISQGYGNTAMIIQKSGNKANITQY 138

RESULT 15
O88HG0
ID Q88HG0 PRELIMINARY; PRT; 157 AA.
AC Q88HG0;
DT 01-JUN-2003 (TrEMBLrel. 24, Created)
DT 01-JUN-2003 (TrEMBLrel. 24, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Curli fiber surface-exposed nuclear CsgB, putative.
GN PP3398.
OS Pseudomonas putida (strain KT2440).
OC Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;
OC Pseudomonadaceae; Pseudomonas.
OX NCBI_TaxID=160488;
RN [1]_
RP SEQUENCE FROM N.A.
RX MEDLINE=22423060; PubMed=12534463;
RA Nelson K.E., Weinel C., Paulsen I.T., Dodson R.J., Hilbert H.,
RA Martins dos Santos V.A.P., Fouts D.E., Gill S.R., Pop M., Holmes M.,
RA Brinkac L., Beanan M., DeBoy R.T., Daugherty S., Kolonay J.,
RA Madupu R., Nelson W., White O., Peterson J., Khouri H., Hance I.,
RA Chris Lee P., Holtzapple E., Scanlan D., Tran K., Moazzez A.,
RA Utterback T., Rizzo M., Lee K., Kosack D., Moestl D., Wedler H.,
RA Lauber J., Stjepandic D., Hoheisel J., Straetz M., Heim S.,
RA Kiewitz C., Eisen J., Timmis K.N., Duesterhoeft A., Tuemmler B.,
RA Fraser C.M.;
RT "Complete genome sequence and comparative analysis of the
RT metabolically versatile Pseudomonas putida KT2440.";
RL Environ. Microbiol. 4:799-808 (2002).
DR EMBL; AB016786; AAN69002.1; -;
DR TIGR; PP3398; -;
DR InterPro; IPR000437; Prok_lipoprot_S.
DR PROSITE; PS00013; PROKAP_LIPOPROTEIN; 1.
KW Complete proteome.
SQ SEQUENCE 157 AA; 16175 MW; C385E98465D2F960 CRC64;

```

Query Match	13.5%	Score 104;	DB 16;	Length 157;
Best Local Similarity	26.4%;	Pred. No. 0.69;		
Matches	37;	Conservative 20;	Mismatches 57;	Indels 26; Gaps 4;
Qy	12	IVVSGSALAGVWPQGGGNGHNGSSGDPDSTLSIYQGSANAALALQSDLKSEFTIT	71	
Db	43	LLPAGAGQAAVIEQQGLG--NRAALDONGALLGRIVCSGGAQEAFTLQ-EGSGLMAIIS	99	
Qy	72	QSGYGRGADYCGQAGDNSTIELTONGFENNATYDQLVTRVVVTHEMAHAGGNNAAALVNQTAS	131	
Db	100	QQCGNGNSASI-----RQSGSSNNAIEQI-----GNDNSASIVQSGS	136	
Qy	132	DSSVMVRQVGFGNNTANQY	151	
Db	137	GLN\$SVTQAGNGQHVQITQY	156	

Search completed: August 2, 2004, 14:54:42  
Job time : 29.7 secs